

GenCore version 5.1.6
Copyright (C) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2004, 11:19:33 ; Search time 128 seconds
(without alignments)
6874.661 Million cell updates/sec

Title: US-10-030-389-12
Perfect score: 1238
Sequence: 1 acatcttggaatggacta.....ttttgtggagggtccagttta 1238

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 355394441 residues
Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/pctus_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	226.2	18.3	1600	2	US-08-933-750C-50
2	226.2	18.3	1600	3	US-09-234-613-50
3	96.2	7.8	471	4	US-09-621-976-1305
4	82.2	5.0	708	4	US-09-270-767-13081
5	62	5.0	8607	4	US-10-204-708-72
6	60.8	4.9	396	4	US-08-640-173-53
7	60.8	4.9	396	4	US-09-713-550-53
8	60.8	4.9	396	4	US-09-825-294-53
9	60.8	4.9	396	4	US-09-970-966-53
10	58	4.7	2447	2	US-09-014-969-14
11	58	4.7	7218	1	US-08-232-463-14
12	57	4.6	240	1	US-08-628-417-6
13	57	4.6	582	4	US-09-787-292-3
14	56.4	4.6	441	4	US-09-601-537-10
15	56.4	4.6	2262	4	US-09-311-021-171
16	56.4	4.6	3275	4	US-09-370-838-151
17	56.4	4.6	3275	4	US-09-854-133-151
18	56.4	4.6	4121	4	US-09-601-537-9
19	56.2	4.5	1798	3	US-09-797-906-1
20	56	4.5	1024	4	US-09-328-475C-50
21	56	4.5	1985	4	US-09-907-794A-212
22	56	4.5	1985	4	US-09-905-125A-212
23	56	4.5	1985	4	US-09-902-775A-212
24	56	4.5	1985	4	US-09-906-700-212
25	56	4.5	1985	4	US-09-903-603A-212
26	55.8	4.5	5152	4	US-10-204-708-47
27	55.4	4.5	1020	4	US-09-328-475C-43

C 28	55.4	4.5	1813	3	US-09-071-224-3	Sequence 3, Appli
C 29	55.4	4.5	6671	1	US-08-280-443-1	Sequence 1, Appli
C 30	55.4	4.5	6671	1	US-08-457-459-1	Sequence 1, Appli
C 31	55.4	4.5	6671	1	US-08-555-678-1	Sequence 1, Appli
C 32	55.4	4.5	6671	5	PCT-US95-02275-1	Sequence 1, Appli
C 33	55.2	4.5	1798	2	US-08-557-128-12	Sequence 12, Appli
C 34	54.6	4.4	249	4	US-09-621-976-1322	Sequence 1322, Ap
C 35	54.6	4.4	396	4	US-09-640-173-33	Sequence 33, Appli
C 36	54.6	4.4	396	4	US-09-713-550-33	Sequence 33, Appli
C 37	54.6	4.4	396	4	US-09-825-294-33	Sequence 33, Appli
C 38	54.6	4.4	396	4	US-09-970-966-33	Sequence 3, Appli
C 39	54.6	4.4	2634	4	US-09-463-238-3	Sequence 1, Appli
C 40	54.4	4.4	1696	4	US-09-835-811-1	Sequence 93, Appli
C 41	54.4	4.4	10144	4	US-10-204-708-93	Sequence 3, Appli
C 42	54.2	4.4	2146	4	US-10-003-392-3	Sequence 3, Appli
C 43	54.2	4.4	2394	4	US-09-800-729-33	Sequence 1, Appli
C 44	54.2	4.4	2674	3	US-09-817-180-1	Sequence 1, Appli
C 45	54.2	4.4	2674	4	US-10-003-395-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-933-750C-50
; Sequence 50: Application US/08933750C
; Patent No. 5932442
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Insite Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/933,750C
FILING DATE: September 23, 1997
CLASSIFICATION: 536
PRIOR APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0356 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 1600 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: U937NOT01

```
; CLONE: 133
US-08-933-750C-50

Query Match      18.3%; Score 226.2; DB 2; Length 1600;
Best Local Similarity 96.7%; Pred. No. 7e-50;
Matches 231; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 ACATCCTTGGATGCGACTAGCAACCAAGGACAGATATTCAGAGTGTACTCTCAAA 60
Db 311 ACATCCTTGGATGCGACTAGCAACCAAGGACAGATATTCAGAGTGTACTCTCAAA 370
QY 61 TCAGTACTCTCAAGCAAGTCCAGAGCCTAGCGTTGCCCACTTAGATCAACAATGGTAG 120
Db 371 TCAGTACTCTCAAGCAAGTCCAGAGCCTAGCGTTGCCCACTTAGATCAACAATGGTAG 430
QY 121 ACCGAGCATCAACTTGTGTTTCTTAAATAATGAAAGGTGAACAGACTAAAGACA 180
Db 431 ACCGAGCATCAACTTGTGTTTCTTAAATAATGAAAGGTGAACAGACTAAAGACA 490
QY 181 AACTGGAACAAGCCCAAAATGAATGAGTGCCTGGAAGTTTACGCTGATAGCCAAACA 239
Db 491 AACTGGAACAAGCCCAAAATGAATGAGTGCCTGGAAGTTTACGCTGATAGCCAAACA 549

RESULT 2
US-09-234-613-50
; Sequence 50, Application US/09234613
; Patent No. 6132973
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/234,613
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,750
; FILING DATE: September 23, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1600 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:

; LIBRARY: U937NOT01
; CLONE: 133
US-09-234-613-50

Query Match      18.3%; Score 226.2; DB 3; Length 1600;
Best Local Similarity 96.7%; Pred. No. 7e-50;
Matches 231; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 ACATCCTTGGATGCGACTAGCAACCAAGGACAGATATTCAGAGTGTACTCTCAAA 60
Db 311 ACATCCTTGGATGCGACTAGCAACCAAGGACAGATATTCAGAGTGTACTCTCAAA 370
QY 61 TCAGTACTCTCAAGCAAGTCCAGAGCCTAGCGTTGCCCACTTAGATCAACAATGGTAG 120
Db 371 TCAGTACTCTCAAGCAAGTCCAGAGCCTAGCGTTGCCCACTTAGATCAACAATGGTAG 430
QY 121 ACCGAGCATCAACTTGTGTTTCTTAAATAATGAAAGGTGAACAGACTAAAGACA 180
Db 431 ACCGAGCATCAACTTGTGTTTCTTAAATAATGAAAGGTGAACAGACTAAAGACA 490
QY 181 AACTGGAACAAGCCCAAAATGAATGAGTGCCTGGAAGTTTACGCTGATAGCCAAACA 239
Db 491 AACTGGAACAAGCCCAAAATGAATGAGTGCCTGGAAGTTTACGCTGATAGCCAAACA 549

RESULT 3
US-09-621-976-1305
; Sequence 1305, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621.976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 1305
; LENGTH: 471
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 299..469
US-09-621-976-1305

Query Match      7.8%; Score 96.2; DB 4; Length 471;
Best Local Similarity 82.7%; Pred. No. 7.9e-16;
Matches 110; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 64 AGTACCTCAAGCAAGTCCAGAGCCTAGCGTTGCCCACTTAGATCAACAATGGTAGACC 123
Db 339 AGGACCCAAAGTCATTTCCAGGTCGGGAGATAAATCAGATCAACAATGGTAGACC 398
QY 124 CAGCGATCAACTTCTTTTCTTAAATAATGAAAGGTGAACAGACTAAAGACAAC 183
Db 399 CAGCGATCAACTTCTTTTCTTAAATAATGAAAGGTGAACAGACTAAAGACAAC 458
QY 184 TGGAAACAAGCCCA 196
Db 459 TGGAAACAAGCCCA 471

RESULT 4
US-09-270-767-13081/c
; Sequence 13081, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
```

```

; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13081
; LENGTH: 708
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-13081

Query Match          5.0%; Score 62.2; DB 4; Length 708;
Best Local Similarity 57.4%; Pred. No. 9.2e-07;
Matches 112; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 963 GTTTTGTAAATTTTTCAGCAAAATTTTATACAGTGCATTAATTTGGGAGAGATAC 1022
DB 657 GTTTTGTAAATTTTTCAGCAAAATTTTATACAGTGCATTAATTTGGGAGAGATAC 598
QY 1023 TGTCAGAAATTAATGCATACATTTTGTACAAATTTGCCATTTTGGGGTGTACGTTTG 1082
DB 597 CCCCCAAAAACATAATCCCGCCCGCCCGGGAATTTTTTTTTTTTTTTTTTTT 538
QY 1083 GTTTTGTAAATTTTTCAGCAAAATTTTATACAGTGCATTAATTTGGGAGAGATAC 1142
DB 537 TTTTGTAAATTTTTCAGCAAAATTTTATACAGTGCATTAATTTGGGAGAGATAC 478
QY 1143 GTTGTGTAAATTTTTCAGCAAAATTTTATACAGTGCATTAATTTGGGAGAGATAC 1157
DB 477 TTTTGTAAATTTTTCAGCAAAATTTTATACAGTGCATTAATTTGGGAGAGATAC 463

RESULT 5
US-10-204-708-72
; Sequence 72, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: BERLIN, Kurt
; APPLICANT: PIEPENBROCK, Christian
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; PRIOR FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 72
; LENGTH: 8607
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-72

Query Match          5.0%; Score 62; DB 4; Length 8607;
Best Local Similarity 51.4%; Pred. No. 3.6e-06;
Matches 143; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 922 TTCAGAAATGGTGCAGTCAAGTGAATGTACAGGTTCACTTTGTAAATTTTCA 981
DB 6825 TTTTAAAGTGTGGGATATAGATGTGATTTATCCGTTTGTAGTTAAAAATATTTAA 6884
QY 982 GCAATTTTATACAGTGTCAATTTAATTTGGGAGAGATACCTCCAGAAAATTAATGCA 1041

; CURRENT APPLICATION NUMBER: US/09640173
; Sequence 53, Application US/09640173
; Patent No. 6613515
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: OVARIAN TUMOR SEQUENCES AND METHODS OF USE THEREFOR
; FILE REFERENCE: 210121.484C2
; CURRENT APPLICATION NUMBER: US/09/640,173
; CURRENT FILING DATE: 2000-08-15
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 53
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(396)
; OTHER INFORMATION: n = A,T,C or G
US-09-640-173-53

Query Match          4.9%; Score 60.8; DB 4; Length 396;
Best Local Similarity 51.2%; Pred. No. 1.6e-06;
Matches 131; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY 959 TTCAGTTTGTAAATTTTTCAGCAAAATTTTATACAGTGCATTTAATTTGGGAGAGG 1018
DB 14 TTTTGTAAATTTTTCAGCAAAATTTTATACAGTGCATTTAATTTGGGAGAGG 73
QY 1019 ATACTGTCCAGAAAATTAATGCATACCTTTGTCACAATTTGCTTTTGGGTGTACGT 1078
DB 74 TTTTGTAAATTTTTCAGCAAAATTTTATACAGTGCATTTAATTTGGGAGAGG 133
QY 1079 TTTGGTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1138
DB 134 TTTTGTAAATTTTTCAGCAAAATTTTATACAGTGCATTTAATTTGGGAGAGG 193
QY 1139 GGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1198
DB 194 TTTTGTAAATTTTTCAGCAAAATTTTATACAGTGCATTTAATTTGGGAGAGG 253
QY 1199 TTGAATGTGTGCAAAAG 1214
DB 254 TCANAAAAAGAAAG 269

RESULT 7
US-09-713-550-53
; Sequence 53, Application US/09713550
; Patent No. 6617109
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND DIAGNOSIS OF OVARIAN CANCER
; OTHER INFORMATION: THERAPY AND DIAGNOSIS OF OVARIAN CANCER
US-09-713-550-53

; CURRENT APPLICATION NUMBER: US/09640173
; Sequence 53, Application US/09640173
; Patent No. 6613515
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: OVARIAN TUMOR SEQUENCES AND METHODS OF USE THEREFOR
; FILE REFERENCE: 210121.484C2
; CURRENT APPLICATION NUMBER: US/09/640,173
; CURRENT FILING DATE: 2000-08-15
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 53
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(396)
; OTHER INFORMATION: n = A,T,C or G
US-09-640-173-53

Query Match          4.9%; Score 60.8; DB 4; Length 396;
Best Local Similarity 51.2%; Pred. No. 1.6e-06;
Matches 131; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY 959 TTCAGTTTGTAAATTTTTCAGCAAAATTTTATACAGTGCATTTAATTTGGGAGAGG 1018
DB 14 TTTTGTAAATTTTTCAGCAAAATTTTATACAGTGCATTTAATTTGGGAGAGG 73
QY 1019 ATACTGTCCAGAAAATTAATGCATACCTTTGTCACAATTTGCTTTTGGGTGTACGT 1078
DB 74 TTTTGTAAATTTTTCAGCAAAATTTTATACAGTGCATTTAATTTGGGAGAGG 133
QY 1079 TTTGGTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1138
DB 134 TTTTGTAAATTTTTCAGCAAAATTTTATACAGTGCATTTAATTTGGGAGAGG 193
QY 1139 GGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1198
DB 194 TTTTGTAAATTTTTCAGCAAAATTTTATACAGTGCATTTAATTTGGGAGAGG 253
QY 1199 TTGAATGTGTGCAAAAG 1214
DB 254 TCANAAAAAGAAAG 269

RESULT 7
US-09-713-550-53
; Sequence 53, Application US/09713550
; Patent No. 6617109
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND DIAGNOSIS OF OVARIAN CANCER
; OTHER INFORMATION: THERAPY AND DIAGNOSIS OF OVARIAN CANCER
US-09-713-550-53
```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2004, 11:19:33 ; Search time 682 Seconds
(without alignments)
9529,008 Million cell updates/sec

Title: US-10-030-389-12

Perfect score: 1238

Sequence: 1 acatccttggaatcgacta.....ttttgtggagggtccagtta 1238

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N Geneseq_23Sep04:*
1: Geneseqn1980s:*
2: Geneseqn1980s:*
3: Geneseqn2000s:*
4: Geneseqn2000s:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2002bs:*
8: Geneseqn2003as:*
9: Geneseqn2003bs:*
10: Geneseqn2003cs:*
11: Geneseqn2003ds:*
12: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1238	100.0	1238	4	AAF31245 Human imm
2	1174.2	94.8	2143	12	ADP07317 Human wt1
3	1116.6	90.2	2772	10	Ades4051 Human pro
4	725.2	58.6	755	3	AAA44466 Human sec
5	296.6	24.0	440	8	ABX50288 Bovine ES
6	249	20.1	322	10	ACD94748 Human col
7	231.4	18.7	2016	5	AA84773 DNA encod
8	226.2	18.3	1600	8	ABX63550 Human cDN
9	226.2	18.3	1622	6	ABL65182 Lung can
10	226.2	18.3	1691	8	ABX63552 Human cDN
11	226.2	18.3	1996	8	ABX63551 Human cDN
12	201.2	16.3	475	6	ABL79439 Human ova
13	184	14.9	1565	12	ADOS8729 DNA encod
14	167.8	13.6	2553	4	ABL10891 Drosophil
15	163.6	13.2	330	4	AA184750 Human pol
16	152.2	12.3	3028	8	AA84774 DNA encod
17	143.4	11.6	3900	8	ABX63451 Human cDN
18	143.4	11.6	3902	10	ADJ56454 Human cDN
19	125.4	10.1	313	12	ADH00046 Kidney cl
20	120.8	9.8	2642	11	ADM03477 Human cDN
21	120.8	9.8	81826	10	ADL13767 Osteoearth

ALIGNMENTS

RESULT 1

AAF31245

ID AAF31245 standard; DNA; 1238 BP.

AC AAF31245;

XX AAF31245;

DT 09-APR-2001 (first entry)

XX Human immune response molecule (IMUN) coding sequence SEQ ID NO: 12.
XX Human; IMUN; immune response molecule; autoimmune disorder;
XX inflammatory disorder; cell proliferation disorder; cancer; ss.

OS Homo sapiens.

PN WO200102569-A2.

XX 11-JAN-2001.

XX 06-JUL-2000; 2000WO-US018505.

XX 06-JUL-1999; 99US-0142572P.

XX 09-SEP-1999; 99US-0153170P.

XX (INCY-) INCYTE GENOMICS INC.

XX Tang YT, Yue H, Yang J, Azimzai Y, Baughn MR, Lu DAM;

XX WPI; 2001-123113/13.

XX P-PSDB; AAB67048.

XX Novel human immune response molecules (IMUN), useful for diagnosing,

XX treating and preventing disorders associated with abnormal expression of

XX IMUN, e.g. Addison's disease, allergies, anemia, asthma and

XX atherosclerosis.

XX Claim 5; Page 90; 95pp; English.

XX The present invention provides the protein and coding sequences for ten

XX human immune response molecules (designated IMUN-1 to IMUN-10). The

XX sequences can be used in the diagnosis, treatment and prevention of

XX autoimmune and inflammatory disorders such as AIDS, atherosclerosis,

XX asthma, allergies, Crohn's disease, multiple sclerosis, irritable bowel

XX syndrome, psoriasis, rheumatoid arthritis and infections, and cell

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

CC proliferation disorders including arteriosclerosis, cirrhosis and cancer
XX
SQ Sequence 1238 BP; 363 A; 249 C; 302 G; 324 T; 0 U; 0 Other;

Db	961		CAGTTT	TGTAATAT	TTTTT	CAGCAAA	TTTTT	TATACAGT	GTCA	TTTAA	TTTTT	TGGGAG	GAGAT	1020
Qy	1021	ACTG	TCAG	AAAAAT	TAAT	TGCAT	ACTTT	TGTC	ACAA	TTTGC	TTTTT	TGTGGG	TGAC	1080
Db	1021	ACTG	TCAG	AAAAAT	TAAT	TGCAT	ACTTT	TGTC	ACAA	TTTGC	TTTTT	TGTGGG	TGAC	1080
Qy	1081	TGGT	TTTTT	TGTT	GTG	TTTTT	TC	TTGG	ATTTT	TTTTT	TC	TTTTT	TAAAG	1140
Db	1081	TGGT	TTTTT	TGTT	GTG	TTTTT	TC	TTGG	ATTTT	TTTTT	TC	TTTTT	TAAAG	1140
Qy	1141	TTG	TG	TTG	TTTTT	TG	CTT	CAAT	CTCT	TCG	CGCG	CTT	CGGAA	1200
Db	1141	TTG	TG	TTG	TTTTT	TG	CTT	CAAT	CTCT	TCG	CGCG	CTT	CGGAA	1200
Qy	1201	GAAT	GTG	CAAA	AGG	GCAT	TTTTT	TG	GAGG	GTCC	GATTA	1238		
Db	1201	GAAT	GTG	CAAA	AGG	GCAT	TTTTT	TG	GAGG	GTCC	GATTA	1238		

RESULT 2	
ADP07317	
ID	ADP07317 standard; DNA; 2143 BP.
XX	
XX	
AC	ADP07317;
XX	
XX	
DT	29-JUL-2004 (first entry)
XX	
XX	
DE	Human Wt1 DNA.
XX	
XX	
KW	ds, proliferative disease; breast; methylation; CpG; bisulfite; human.
XX	
XX	
OS	Homo sapiens.
XX	
PN	DE10255104-A1.
XX	
XX	
PD	11-MAR-2004.
XX	
PF	26-NOV-2002; 2002DE-01055104.
XX	
PR	27-AUG-2002; 2002DE-01039313.
XX	
PA	(EPIC-) EPIGENOMICS AG.
XX	
PI	Maier S;
XX	
DR	WPI; 2004-284340/27.
DR	GENBANK; NW_004906.
XX	
XX	
PT	Analyzing proliferative diseases of breast cells, useful e.g. for
PT	diagnosis, prognosis and treatment of breast cancer, by determining
PT	methylation status of specific genes.
XX	

the oligomer, optionally with extension of the hybridised oligomer by at least one base, or detection is by sequencing. The amplification may use methylation-specific primers. Alternatively, genomic DNA is extracted from a sample and digested with methylation-specific restriction enzymes, then the digestion fragments detected, optionally after amplification. In either method, more than 10 fragments of 100-200 bp are amplified in a single reaction vessel, using a heat-resistant DNA polymerase in PCR. The amplicons carry detectable markers, e.g. fluorophores, radioisotopes and/or releasable fragments of known mass that can be detected by mass spectrometry. The method is used for characterisation, classification, differentiation, staging, phase-estimation, diagnosis and/or therapy of proliferative diseases of breast cells. The method provides very specific classification of proliferative diseases, allowing better treatment. It can both characterise methylation status and detect single-nucleotide polymorphisms. This sequence represents human gene used to illustrate the method of the invention. NOTE: This sequence does not appear in the printed specification but has been retrieved from Genbank.

Sequence 2143 BP: 619 A; 431 C; 498 G; 595 T; 0 U; 0 Other;

Query Match	94.8%	Score 1174.2;	DB 12;	Length 2143;
Best Local Similarity	97.8%	Pred. NO. 1.6e-262;		
Matches 1212; Conservative	0;	Mismatches 23;	Indels 4;	Gaps 2;

QY	1	ACATCTTGGAATGCGACTAGCAACCAAGGACAAGATATTCAGAGTGTACTACTC	60
DB	443	ACATCTTGGAATGCGACTAGCAACCAAGGACAAGATATTCAGAGTGTACTACTC	502
QY	61	TCCAGTACCTCAAAGCAAGTCCAGCAGCGCTAGCGTTGCCAACTTAGATCAACAATGGTAG	120
DB	503	TCCAGTACCTCAAAGCAAGTCCAGCAGCGCGTGTGCCAACTGAGATCAACAATGGTAG	562
QY	121	ACCCAGGATCAACTGTTTCTTAAATGAAGTGAACCTGGAACAGACTGAAGACA	180
DB	563	ACCCAGCGATCAACTGTTTCTTAAATGAAGTGAACCTGGAACAGACTGAAGACA	622
QY	181	AACTGGAACAAGCCCAAAATGAACTGAGTGCCTTGGAAAGTTTACGCCTGATAGCCAAACAG	240
DB	623	AACTGGAACAAGCCCAAAATGAACTGAGTGCCTTGGAAAGTTTACGCCTGATAGCCAAACAG	682
QY	241	GGAAAAAGTTAATGGCGAAATGTGCGAATGCTTATCCAGGAGATCAAGACTTGGAGGC	300
DB	683	GGAAAAAGTTAATGGCGAAATGTGCGAATGCTTATCCAGGAGATCAAGACTTGGAGGC	742
QY	301	AGCTGTCCCGGGACGTATTGCACAACTTCAAGCAGAGTTGGCTTTACAGAGAAATACA	360
DB	743	AGCTGTCCCGGGACGTATTGCACAACTTCAAGCAGAGTTGGCTTTACAGAGAAATACA	802
QY	361	GTGAGGAGCTTAAAGCAGTCAGATGAACTGAATGACTTCATCTCCAGCTTGATGAAG	420
DB	803	GTGAGGAGCTTAAAGCAGTCAGATGAACTGAATGACTTCATCTCCAGCTTGATGAAG	862
QY	421	AAGTAGAGGGTATCCAGAGTACCATTCTAGTTCTGCAGCAGCAGCTGAAGCAGACGCGC	480
DB	863	AAGTAGAGGGTATCCAGAGTACCATTCTAGTTCTGCAGCAGCAGCTGAAGCAGACGCGC	922
QY	481	AGCAGTTGGCTCAGTACCAGCAGCAGCTCTCAGGCCTCTGCCCCAAGTACCAGCAGGA	540
DB	923	AGCAGTTGGCTCAGTACCAGCAGCAGCTCTCAGGCCTCTGCCCCAAGTACCAGCAGGA	982
QY	541	CTACAGCTTCTGAACCTGTAGAAACAGTCAGAGGCCCAAGTAAAGCTGCAGTCGCTGA	600
DB	983	CTACAGCTTCTGAACCTGTAGAAACAGTCAGAGGCCCAAGTAAAGCTGCAGTCGCTGA	1042
QY	601	CAAAACGGAACCAAGTAATGGTAGCTCCTCCCGCCAGAGAGCGTCTGGGTCTGGATTTCACA	660
DB	1043	CAAAACGGAACCAAGTAATGGTAGCTCCTCCCGCCAGAGAGCGTCTGGGTCTGGATTTCACA	1102
QY	661	GGGAGGGCAACACAAACGGAAGATGACTTCCCTTCTCCAGGGAATGGTAATAGTCCT	720
DB	1103	GGGAGGGCAACACAAACGGAAGATGACTTCCCTTCTCCAGGGAATGGTAATAGTCCT	1162
QY	721	CCAAACAGCTCAGAGAGAGAACTCGGCAGAGAGGAGTAGTGGTTTACGTAAATCAACTCAGTG	780

Db	1163	CCAA	CAGCTCAGAGGAGAGAACTGGCAGAGGAGGTAGTGGTTACGTTAAATCAACTCAGTG	1222
Qy	781	CGGG	TATGAAAGTGTAGACTCTCCACGGGCAGTGAATACTCTCTCACACACCAATCAA	840
Db	1223	CGGG	TATGAAAGTGTAGACTCTCCACGGGCAGTGAATACTCTCTCACACCAATCAA	1282
Qy	841	ATGAC	CACAGACTCCAGTCAATGACCCCTCAAGAGAGAGAAAGCAGTGAGTGGGAAAGGTAATC	900
Db	1283	ATGAC	CACAGACTCCAGTCAATGACCCCTCAAGAGAGAGAAAGCAGTGAGTGGGAAAGGTAATC	1342
Qy	901	GAAC	TGGGGTTCGGCCACGTTTCAGAAATGGCTTGGACTCAAGTGTAAATGTACAGGGTT	960
Db	1343	GAAC	TGGGGTTCGGCCACGTTTCAGAAATGGCTTGGACTCAAGTGTAAATGTACAGGGTT	1402
Qy	961	CAGT	TTTGTAAATATTTTTTCAGAAATTTTTATACAGTGCATTTAAATTTGGGAGAGGAT	1020
Db	1403	CAGT	TTTGTAAATATTTTTTCAGAAATTTTTATACAGTGCATTTAAATTTGGGAGAGGAT	1462
Qy	1021	ACTG	TCAGAAATTAATGCATCTATTGTGCACAAATTTGGCTTTTCTGGGTGTACGGTT	1080
Db	1463	ACTG	TCAGAAATTAATGCATCTATTGTGCACAAATTTGGCTTTTCTGGGTGTACGGTT	1522
Qy	1081	TGGT	TTTTTTTTTGTGTGTTTTTTTCTTTGGAATTTTTTTCTTTTCTTTTTTAAGAGGTGG	1140
Db	1523	TGGT	TTTTTTTTTGTGTGTTTTTTTCTTTGGAATTTTTTTCTTTTCTTTTTTTT---	1579
Qy	1141	TTGT	TGTTTTTTTGCTTCAATACCTTCGCGCGTTGGGAATGTGAACAGTTAAATTACTTT	1200
Db	1580	TTTT	TTTTTTTTTTTGCTTCAATACCTTCGCGCGTTGGGAATGTGAACAGTTAAATTACTTT	1639
Qy	1201	GAAT	TGTTGC-AAAAGGOCATTTTGTGGAGGGTCAGTTA	1238
Db	1640	GAAT	TGTTGCTAAAGAGCATTTTGTGTAGGGTCAAGTTA	1678

RESULT 3
ADE54051

AD54051
ID AD54051 standard: cDNA: 2772 bp.

AA ADE54051;
AC

DT 29-JAN-2004 (first entry)

XX
DE Human prostate cancer cDNA #398.

XX Human; prostate cancer; ss; cDNA combination; differential expression;
KW gene.
KW

XX Homo sapiens.

XX
PN
IIS2003190640-A1.XX
09-00T-2003

X

XX

XX
XX

PA (FARL//) FARIS M.
PA (PEAR//) PEARSON

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

XX

XX
XX

PT progression of treatment of prostate cancer.

XX
CC
CL-10 1. SEC IN NO 398. 42mm. Enclish

XX
... a number of CDNA

expressed in prostate cancer. The invention also relates to a method for detecting differential expression of one or more cDNAs in a sample containing nucleic acids by hybridising a substrate with the nucleic acids, thus forming one or more hybridisation complexes, detecting hybridisation complex formation and comparing the complexes formed with standard complexes, where differences between the standard and the sample complex formation indicate differences between the standard and the sample. The differential expression is diagnostic of cDNAs in the sample. The invention also relates to proteins and antibodies related to the cDNAs. The combination is useful for diagnosing, treating or monitoring the progression of treatment of prostate cancer. The antibodies are useful for detecting prostate cancer. This sequence represents a human prostate cancer cDNA of the invention.

Sequence 2772 BP; 740 A; 680 C; 654 G; 688 T; 0 U; 0 Other;

Query Match 90.2%; Score 1116.6; DB 10; Length 2772;
Best Local Similarity 99.2%; Pred No. 4.1e-249;
Matches 1122; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 ACATCCTTGGAAATGCGACTAGCAACCAAGGACAGATATTCAAGAGTGTACTACTCAA 60
DB 1641 ACATCCTTGGAAATGCGACTAGCAACCAAGGACAGATATTCAAGAGTGTACTACTCAA 1700
QY 61 TCCAGTACCTCAGCAAGTCCAGCAGCTAGCGTTGCCCACTTAGATCAACAATGGTAG 120
DB 1701 TCCAGTACCTCAGCAAGTCCAGCAGCTAGCGTTGCCCACTTAGATCAACAATGGTAG 1760
QY 121 ACCCAGCGATCAACTGTTTTCTTAAATGAAAGGTGAAGTGAAGCAAGTGAAGCA 180
DB 1761 ACCCAGCGATCAACTGTTTTCTTAAATGAAAGGTGAAGTGAAGCAAGTGAAGCA 1820
QY 181 AACTGGAAACAGCCAAATGAATGAATGCTTATCCAGGAGATCAAGAGCTTGAAGCA 240
DB 1821 AACTGGAAACAGCCAAATGAATGAATGCTTATCCAGGAGATCAAGAGCTTGAAGCA 1880
QY 241 GGAAGAAAGTAAATGGCAAGTGTGCAATGCTTATCCAGGAGATCAAGAGCTTGAAGCA 300
DB 1881 GGAAGAAAGTAAATGGCAAGTGTGCAATGCTTATCCAGGAGATCAAGAGCTTGAAGCA 1940
QY 301 AGCTGTCCAGGAGCGTATTGCAAACTTGAAGCAGAGTGGCTTTACAGAAGAAATACA 360
DB 1941 AGCTGTCCAGGAGCGTATTGCAAACTTGAAGCAGAGTGGCTTTACAGAAGAAATACA 2000
QY 361 GTGAGGAGCTTAAAGCAGTCAAGATCACTGAATGCTTATCCAGGAGATCAAGAGCTTGAAG 420
DB 2001 GTGAGGAGCTTAAAGCAGTCAAGATCACTGAATGCTTATCCAGGAGATCAAGAGCTTGAAG 2060
QY 421 AAGTAGAGGCTATCCAGATACCAATCTAGTTCTGCAAGCAGAGCTGAAGGAGACAGCC 480
DB 2061 AAGTAGAGGCTATCCAGATACCAATCTAGTTCTGCAAGCAGAGCTGAAGGAGACAGCC 2120
QY 481 AGCAGTTGGCTCAGTACCAAGCAGCAGTCTAGGCTCTGCCCCAAGTACCAGCAGGA 540
DB 2121 AGCAGTTGGCTCAGTACCAAGCAGCAGTCTAGGCTCTGCCCCAAGTACCAGCAGGA 2180
QY 541 CTACAGCTTCTGAACTGTAGAAAGTCAAGTCAAGGAGCAGAGTCAAGTCAAGTCA 600
DB 2181 CTACAGCTTCTGAACTGTAGAAAGTCAAGTCAAGGAGCAGAGTCAAGTCAAGTCA 2240
QY 601 CAAACGACCAAGTAATGGTAGCTCTCCCGCAGAGGAGCTGGGTCTGGAATTTACA 660
DB 2241 CAAACGACCAAGTAATGGTAGCTCTCCCGCAGAGGAGCTGGGTCTGGAATTTACA 2300
QY 661 GGGAGGGCAACACAAACGAAGATGACTTCTCTTCTCCAGGAGATGTAAGTCTCT 720
DB 2301 GGGAGGGCAACACAAACGAAGATGACTTCTCTTCTCCAGGAGATGTAAGTCTCT 780
QY 721 CCAACAGCTCAGAGGAGAACTGGCAGAGAGAGTGTAGTACGTAATCAACTCAGTG 780
DB 2361 CCAACAGCTCAGAGGAGAACTGGCAGAGAGAGTGTAGTACGTAATCAACTCAGTG 2420
QY 781 CGGGGTATGAAGTGTAGACTCTCCCAAGGAGCAGTGAAGTCTCTCAGACCAATCAA 840

DB 2421 CGGGGTATGAAGTGTAGACTCTCCACGGGAGTGAAAACCTCTCTCACACCAATCAA 2480
QY 841 ATGACACAGACTCCAGTCTATGACCCCTCAAGAGAGAAAGCAGTGTAGTGGAAAGTAAATC 900
DB 2481 ATGACACAGACTCCAGTCTATGACCCCTCAAGAGAGAAAGCAGTGTAGTGGAAAGTAAATC 2540
QY 901 GAACTGTGGGTTCCCGCCACAGTTCAGAAATGGCTTGGACTCAAGTCAATGTAAATGTACAGGGTT 960
DB 2541 GAACTGTGGGTTCCCGCCACAGTTCAGAAATGGCTTGGACTCAAGTCAATGTAAATGTACAGGGTT 2600
QY 961 CAGTTTGTAAATATTTTTCAGCAAAATTTTATACAGTGTCAATTTAAATTTGGGAGAGAT 1020
DB 2601 CAGTTTGTAAATATTTTTCAGCAAAATTTTATACAGTGTCAATTTAAATTTGGGAGAGAT 2660
QY 1021 ACTGTCCAGAAATTAATGACATCTTTGTCAAAATTTGCCCTTTTGTGGGTGTACGTTT 1080
DB 2661 ACTGTCCAGAAATTAATGACATCTTTGTCAAAATTTGCCCTTTTGTGGGTGTACGTTT 2720
QY 1081 TGGTTTTTTTGTGTTTCTTTTCTTGGATTTTCTTTTCTTTTCTTTTCTTTT 1131
DB 2721 TGGTTTTTTTGTGTTTCTTTTCTTGGATTTTCTTTTCTTTTCTTTTCTTTT 2771

RESULT 4

AAA44466

ID AAA44466 standard; cDNA; 755 BP.

XX AC AAA44466;

XX DT 21-AUG-2000 (first entry)

XX DE Human secreted expressed sequence tag SEQ ID NO:1041.

Human; mouse; chicken; rat; secreted expressed sequence tag; sEST;
expressed sequence tag; EST; probe; chemotactic; proliferative;
immunomodulatory; hematopoietic; chemokinetic; analgesic; haemostatic;
thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
antiviral; antidiabetic; antisthmatic; vulnerary; antiparkinsonian;
antitumor; osteopathic; neuroprotective; nootropic; antipsoriatic;
cerebroprotective; anticonvulsant; antidepressant; gene therapy; vaccine;
autoimmune disorder; multiple sclerosis; allergic condition;
insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
central nervous system disorder; Alzheimer's disease; stroke;
Parkinson's disease; Huntington's disease; coagulation disorder;
haemophilia; thrombosis; inflammatory disorder; Crohn's disease; tumour;
infection; depression; psoriasis; ss.

XX OS Homo sapiens.

XX PN WO200021991-A1.

XX PD 20-APR-2000.

XX PF 15-OCT-1999; 99WO-US024206.

XX PR 15-OCT-1998; 98US-010436P.

XX PA (GENY) GENETICS INST INC.

XX PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;

XX PI Merberg D, Treacy M, Bowman MR;

XX DR WPI; 2000-317938/27.

Isolated polynucleotides, and encoded proteins, comprising secreted
expressed sequence tags (sESTs), useful for treating various disorders
such as autoimmune, infectious, and central nervous system disorders.
Claim 1; Page 470-471; 803pp; English.

XX AAA443426 to AAA45925 represent specifically claimed secreted expressed


```

CC seqdata.uspto.gov/sequence.html?DocID=20020137160
XX Sequence 440 BP; 123 A; 85 C; 92 G; 140 T; 0 U; 0 Other;
XX
Query Match      24.0%; Score 296.6; DB 8; Length 440;
Best Local Similarity 89.5%; Pred. No. 6.2e-59;
Matches 332; Conservative 0; Mismatches 34; Indels 5; Gaps 1;
QY 742 CTGGCAGAGGAGGTAGTGGTTACGTPAAATCAACTCAGTCGCGGGTATGAAAGTGTAGACT 801
Db 1 CTGGCAGAGGAGGTAGTGGTTACGTPAAATCAACTCAGTCGCGGGTATGAAAGTGTAGACT 60
QY 802 CTCACCGCGGAGTGAAGAACTCTCTCACACACCAATCAATACACAGACTCCAGTCATG 861
Db 61 CTCACCGCGGAGTGAAGAACTCTCTCACACACCAATCAATACACAGACTCCAGTCATG 120
QY 862 ACCCTCAAGAGAGAGAAAGCAGTGTGGTGGAAAGGTAAATCGAACTGTGGTTCCCGCCACG 921
Db 121 ACCCGCAAGAGAGAGAAACCTGTGAGTGGAAAGGTAAACCGAACTGTGGTTCCCGCCACG 180
QY 922 TTCAGATGGCTTGGACTCAAGTGTAAATGTACAGGGTTCAGTTTGT-----AATATT 976
Db 181 TTCAGATGGCTTGGACTCAAGTGTAAATGTACAGGGTTCAGTTTGTAAAAAATAAAT 240
QY 977 TTTCAGCAAAATTTTATACAGTGTCAATTAATTTGGAGAGGACTACTCCAGAAATTA 1036
Db 241 TTTCAGCAAAATTTTATACAGTGTCAATTAATTTGGAGAGGACTACTCCAGTAAATTA 300
QY 1037 ATGATACCTTTTGTACAAATTTGCTTTTGTGGGTGTACGTTTGTGTTTTTTTGTG 1096
Db 301 ATGATACCTTTTGTACAAATTTGCTTTTGTGGGTGTACGTTTGTGTTTTTTTGTG 360
QY 1097 TTTTCTTTCTT 1107
Db 361 TTTCTTTGCTT 371
RESULT 6
ID ACD94748 standard; cdna; 322 BP.
XX ACD94748;
XX
XX 23-SEP-2003 (first entry)
XX Human colon cancer cell expressed cdna #3160.
XX
XX Open reading frame detection; genome sequencing; colon cancer;
KW breast cancer; population genome analysis; genetic shift; cancer;
KW antibiotic resistance; antibiotic non-tolerance; congenital disease;
KW agriculture; food crop genome; resistance gene; retrovirus;
KW influenza virus; eukaryotic pathogen detection; trypanosome; Plasmodium;
KW gene; ss.
XX
XX Homo sapiens.
XX
XX US2002155438-A1.
XX
XX 24-OCT-2002.
XX
XX 27-SEP-1999; 99US-00406117.
XX
XX 20-NOV-1998; 98US-00196716.
XX
XX (SIMP/) SIMPSON A J G.
XX (NETO/) NETO E D.
XX (BREN/) BRENTANI R R.
XX
XX Simpson AJG, Neto ED, Brentani RR;
XX
XX WPI; 2003-182626/19.
XX
XX Determining open reading frames of genome of an organism e.g. a human

```

```

PT suffering from cancer involves use of single oligonucleotide primer at
PT low stringency for preparing single-stranded cDNA from mRNA of
PT individual.
XX
XX Example 9; Page 465; 959pp; English.
XX
XX The invention describes a method of determining open reading frames in
XX the genome of organism, comprising contacting mRNA from cell of organism
XX with a single oligonucleotide primer (I) at low stringency, preparing
XX single-stranded cDNA by reverse transcribing mRNA with (I), amplifying
XX cDNA, sequencing the product, and repeating the contacting and sequencing
XX and amplifying steps with different primers and sequencing, resulting
XX nucleic acids. The method is useful for: determining that a known
XX nucleotide sequence from a genome of an organism corresponds to a
XX nucleic acid molecule from a genome of an organism; and for sequencing
XX all or part of a genome of an organism. mRNA is obtained from mammalian
XX or human cell which is associated with a pathological condition e.g. a
XX colon cancer or breast cancer cell. The method is useful for analyses of
XX populations of subjects and can be used to carry out genetic analyses of
XX large or small populations. further, it can be used to study living
XX systems to determine if, e.g. there have been genetic shifts which render
XX an individual or population more or less likely to be afflicted with
XX diseases such as cancer, to determine antibiotic resistance or non-
XX tolerance, and so forth. The method can also be used in the study of
XX congenital diseases, and the risk of affliction to a fetus, as well as
XX the study of whether the conditions are likely to be passed to offspring
XX through ova or sperm. The analyses for pathological conditions can be
XX carried out in all animals, plants, birds, fish, etc. Using this method,
XX in the area of agriculture, for example the genomes of food crops can be
XX studied to determine if resistance genes are present, defects in plant
XX genomes can also be studied in this way. Similarly, the method permits
XX determination of the pathogens which integrate into the genome, such as
XX retroviruses and other integrating viruses such as influenza virus, have
XX undergone shifts or mutations, which may require different approaches to
XX therapy. This method is also applied to eukaryotic pathogens, such as
XX trypanosomes, different types of Plasmodium, etc. The method essentially
XX eliminates sequencing of non-coding portions. This sequence represents a
XX polynucleotide isolated from human colon cancer cell cdna library
XX
XX Sequence 322 BP; 63 A; 76 C; 73 G; 106 T; 0 U; 4 Other;
XX
Query Match      20.1%; Score 249; DB 10; Length 322;
Best Local Similarity 94.7%; Pred. No. 6.3e-48;
Matches 266; Conservative 0; Mismatches 14; Indels 1; Gaps 1;
QY 2 CATCCTTGGATGGACTAGCAACCAAGGGGACAGAGATATTCAAGAGTGTACTTCAAAAT 61
Db 280 CATCCTTGGATGGACTAGCAATNCAA-GGTCAAGAGATGCAAGAGTGTACTTCAAAAT 222
QY 62 CCAGTACCTCAAGCAAGTCCAGCAGCCTAGCGTTGCCCACTTAGATCAACAATGGTAGA 121
Db 221 CCAGTACCTCAAGCAAGTCCAGCAGCAGCGGTTGNCATCTGAGATCAACAATGGTAGT 162
QY 122 CCCAGCGATCAACTGTGTTTTTCTTAAATAATGAAAGGTGAACTGGAACAGACTAAAGCAA 181
Db 161 CCCAGCGATCAACTGTGTTTTTCTTAAATAATGAAAGGTGAACTGGAACAGACTAAAGCAA 102
QY 182 ACTGGACCAAGCCCAAAATGAACTGAGTGCCTGGAAGTTTACGCTTATAGCCAAACAGG 241
Db 101 ACTGGACCAAGCCCAAAATGAACTGAGTGCCTGGAAGTTTACGCTTATAGCCAAACAGG 42
QY 242 GAAAAAGTTAATGGCGAAGTGTGCAATGCTTATCCAGGAGA 282
Db 41 GAAAAAGTTAATGGCGAAGTGTGCAATGCTTATCCAGGAGA 1
XX
RESULT 7
AAS84773
ID AAS84773 standard; cdna; 2016 BP.
XX
XX AAS84773;
XX

```

DT XX 13-FEB-2002 (first entry)
DE XX DNA encoding novel human diagnostic protein #20577.
KW Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder; ss.
OS Homo sapiens.
XX W0200175067-A2.
PN 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US008631.
XX 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
PA Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
DR P-PSDB; ABG20586.
XX New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.
XX Claim 1; SEQ ID NO 20577; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activities. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
SQ Sequence 2016 BP; 403 A; 327 C; 779 G; 501 T; 0 U; 0 Other;
Query Match 18.7%; Score 231.4; DB 5; Length 2016;
Best Local Similarity 99.6%; Pred. No. 1.4e-43;
Matches 232; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 387 GAACCTGAATGATCTTCATCCAGCTTGATGAAGAGTAGAGGTATGCAGAGTACCATT 446
DB 1141 GAACCTGAATGATCTTCATCCAGCTTGATGAAGAGTAGAGGTATGCAGAGTACCATT 1200
QY 447 CTAGTCTCGACGACCTGAAGAGACGACGACGAGTGGCTGAGTACACGACGAG 506
DB 1201 CTAGTCTCGACGACCTGAAGAGACGACGACGAGTGGCTGAGTACACGACGAG 1260
QY 507 CAGTCTCAGGGCTCTGCCCAAGTACACGACGAGGACTACAGCTTCTGAACCTGTAGACAG 566
DB 1261 CAGTCTCAGGGCTCTGCCCAAGTACACGACGAGGACTACAGCTTCTGAACCTGTAGACAG 1320
QY 567 TCAGAGGCCACCAAGTAAAGACTGCGTCTGACAAACGACCAAGTATGG 619

DB 1321 TCAGAGGCCACCAAGTAAAGACTGCGTCTGACAAACGACCAAGTATGG 1373
RESULT 8
ID ABX63550 standard; cDNA; 1600 BP.
XX AC ABX63550;
XX DT 26-FEB-2003 (first entry)
XX DE Human cDNA #550 differentially expressed in activated vascular tissue.
XX Human; gene; ss; vascular tissue; cytostatic; atherosclerosis; cardiatic; hypotensive; antidiabetic; gynaecological; vasotropic; cerebroprotective; gene therapy; vascular disease; cancer; coronary; artery disease; hypertension; diabetes; pre-eclampsia; restenosis;
KW ischaemia-reperfusion injury; stroke.
XX OS Homo sapiens.
XX US2002137081-A1.
XX PD 26-SEP-2002.
XX PF 08-JAN-2002; 2002US-00044090.
XX PR 28-JUL-2000; 2000US-0222469P.
XX PR 08-JAN-2001; 2001US-0260483P.
XX (BAND/) BANDMAN O.
XX Bandman O;
XX WPI; 2003-110597/10.
XX Combination for diagnosing, staging, treating, or monitoring the progression of treatment of a vascular disease, e.g. atherosclerosis, comprises several cDNAs that are differentially expressed in activated vascular tissue.
XX Claim 1; Page; 18pp; English.
XX This invention relates to a combination comprising several cDNAs that are differentially expressed in activated vascular tissue. The invention also discloses a high throughput method for detecting differentially expressed cDNAs in a sample. The cDNAs of the invention may have antiatherosclerotic; cytostatic; cardiatic; hypotensive; antidiabetic; gynaecological; vasotropic and cerebroprotective activities and may be used in gene therapy. The cDNAs of the invention may be used in a high-throughput methods for detecting differential expression of one or more cDNAs in a sample, or screening several molecules or compounds to identify a molecule or compound that specifically binds a cDNA of the invention. A protein encoded by the cDNA may be used to screen several molecules or compounds to identify a ligand that specifically binds to the protein, or to produce or purify an antibody to the protein that can be used to detect a protein in a sample or purify a natural or recombinant protein from a sample. The nucleotides may be useful for diagnosing, staging, treating, or monitoring the progression of treatment of a vascular disease, e.g. atherosclerosis, cancer, coronary artery disease, hypertension, diabetes, pre-eclampsia, ischaemia-reperfusion injury, restenosis, or stroke. The cDNAs can also be used for large-scale genetic or gene expression analysis of several new nucleic acid molecules. Antibodies to the proteins encoded by the cDNAs are useful for diagnosing pre-pathologic disorders, and chronic or acute diseases associated with abnormalities in the expression, amount or distribution of the protein. The present sequence represents a cDNA of the invention that is differentially expressed in activated vascular tissue. Note: The sequence data for this patent did not form part of the specification, but was obtained in electronic format directly from USPTO at http.segdata.uspto.gov/sequence.html?docID=20020137081

(PEKE) PE CORP NY.
PA PA
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX PT WPI; 2001-656860/75.
XX DR P-PDSB; ABB66788.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX Claim 1; SEQ ID NO 27155; 2ipp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed RNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC AB372072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ffp.wipo.int/pub/published_pct_sequences
XX sequence 2553 BP; 783 A; 641 C; 624 G; 505 T; 0 U; 0 Other;
SQ

	Matches	299;	Conservative	0;	Mismatches	217;	Indels	0;	Gaps	0;
Qy	1	ACA	TCTTGG	ANTGC	AGTAGCA	ACCAAGG	CAAGATATTCA	AGAGTGTACT	ACTCAA	60
Db	821	AGAT	TCTCTA	TGAC	GGCTGG	CCAA	CAAGG	CAAGATNTC	AGGACTAT	880
Qy	61	TCCAG	TACCTCA	AGCAAG	TCCAG	CGCTAG	CGTTGCC	CAACTTAG	ATCAACAATCG	120
Db	881	TTGCC	NGATAC	AGGCCCA	AGCGGCC	CAACTGCC	CTCGCAT	TGCGCACT	TGCTCTCTCG	940
Qy	121	ACCAG	CGATCA	ACTG	TTTTCT	TCTAAA	ATGAA	AGTGAA	CTGGAAC	180
Db	941	ATCTG	GGGTCA	ATCTCT	CTTTCG	AGCGGT	CTAAG	AAGGA	ACTTAAGG	1000
Qy	181	AAC	TGGAAC	AAGCC	CAAAATGA	ACTGAT	CGTCC	TGGAA	GTTCGCTGAT	240
Db	1001	AGCT	TGAGG	AGACCC	AGAC	AGCTG	TCGCAT	GAGATTC	AGCCG	1060
Qy	241	GGAAAA	AGTTAA	TGGCG	GAAGTGT	CGAATG	CTTTAT	CCAGG	AGATCAAG	300
Db	1061	GC	AACCCCT	TAATGG	CAATGG	CCAGTGC	TCTCTAT	CAGG	AAACAGG	1120

QY	301	AGCTGTCCCGAGGACGTTATTGCACAACCTTGAACGACAGTTGGCTTTTACAGAAAGAAATACA	360
DB	1121	TGACCTCCAAACGGGAGACTGGCCAGAGCTGGAGACCGAACTGGGCATGCGAGACAGCTTCA	1180
QY	361	GTGAGGAGCTTTAAAAGCAGTCAGGATGAACCTGAATGACTTCATCATCCAGCTTGTATGAAG	420
DB	1181	GCGAAGAGGTCAGAAATCGCAGTCGGAGTTGCACGACTTCCTGCAGGAGCTGGACGAGG	1240
QY	421	AAGTAGAGGGTATGCAGAGTACCATTCCTAGTTCTGCAGCAGCAGCTGAGGACACAGCC	480
DB	1241	ATGTGGAGGGCATGCAGAGCACCATTCTGTTTTGCAGCAAGAACTAAAGACCACACGG	1300
QY	481	AGCAGTTGGCTCAGTACACGACGACGAGCTCTCAG	515
DB	1301	ATCCCATACAAACGCTTGAGAGGAGAACGCCACG	1335

RESULT 15
AAI84750/c
ID AAI84750 standard; cDNA; 330 BP.

QY 1195 TACTTTGAATGTTGCAAAAGGGCAATTT 1222
Db 31 TACTTTGAATGTTGCTAAAAGGACATTT 4

Search completed: November 14, 2004, 13:05:30
Job time : 687 secs

XX 06-NOV-2001 (first entry)
XX Human polynucleotide SEQ ID NO 4810.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation, ss.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US004927.
XX
PR 28-FEB-2000; 2000US-00515126.
PR 18-MAY-2000; 2000US-00577409.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-514838/56.
DR P-PSDB; AAO04819.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing diagnosing
PT and treating e.g. leukemia, inflammation and immune disorders.
XX
PS Claim 1; SEQ ID NO 4810; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pat_sequences
XX
SQ Sequence 330 BP; 151 A; 63 C; 49 G; 67 T; 0 U; 0 Other;

Query Match 13.2%; Score 163.6; DB 4; Length 330;
Best Local Similarity 70.7%; Pred. No. 4.3e-28;
Matches 232; Conservative 0; Mismatches 94; Indels 2; Gaps 1;

QY 895 GTAATCGAAGTGGGTCCGGCCAGCTTCAGAAATGCTTGAAGTAAATGATAC 954
Db 329 GAAATTAACCTGGGTTCCTCCGCCCTTTAAAGAGGTTGGCCCAAGTTAAAAACC 270

QY 955 AGGTTTCAGTTTGTATATTTTTCAGCAAAATTTTATACATGTCATTTTGGGA 1014
Db 269 CGGGGCCCGTTTTTAAAAATTTTTCCTCCCAATTTTTCAGGGTCTTTTGGGG 210

QY 1015 GAGGATACCTCCAGAAAAATTAATGATACATTTTGTCAAAATTTGCTTTTGTGGGTG 1074
Db 209 AAGGAATCTTTCCAAAAAAGAAATTTTTCCTCCATTCCTTTTGGGGGG 150

QY 1075 ACCTTTTGGTTTTTTTGTGTTTTTTTCTTTGGATTTTTTTCTTTCTTTTAAAG 1134
Db 149 GGGGTTTGGGTTTTTTTGGGGTTTTTTTGGGTTTTTTTGGGTTTTTTT--T 92

QY 1135 AGGTGGTGTGTTTTTTTCTTTCAATACCTTCGCGGCTTGGGAATTTGAACAGTTAAT 1194
Db 91 TTTTTTTTTTTTTTTGTGCTTCATCTCTCGCGCTTTGGAATTTGAACAGTTAAT 32

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2004, 11:19:33 ; Search time 5521 Seconds
(without alignments)
10603.993 Million cell updates/sec

Title: US-10-030-389-12
Perfect score: 1238
Sequence: 1 acatcttggaatgcagcta.....ttttgtggagggtccagtta 1238

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_btg.*
- 3: gb_in.*
- 4: gb_cm.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_rc.*
- 11: gb_sts.*
- 12: gb_ey.*
- 13: gb_un.*
- 14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1238	100.0	1238	6	AX069004 Sequence
2	1174.6	94.9	2108	9	BC000383 Homo sapi
3	1159.8	93.7	2134	9	AF374416 Homo sapi
4	980	79.2	1802	10	MMU276707 Mus muscu
5	953.8	77.0	2291	6	C0845798 Sequence
6	953.8	77.0	2291	9	AX131393 Homo sapi
7	945.8	76.2	1167	9	HA2276706 Homo sapi
8	797.8	64.4	165802	9	AL135914 Human DNA
9	666.6	53.8	203077	10	AC127172 Mus muscu
10	666.6	53.8	203949	10	AL589878 Mouse DNA
11	654.6	52.9	110000	2	AC074222.1 of
12	654.6	52.9	110000	2	Continuation (3 of
13	654	52.8	1888	5	BX950648 Gallus ga
14	608.6	49.2	165096	2	AC112423 Rattus no
15	608.6	49.2	246585	2	AC108954 Rattus no
16	590.8	47.7	2179	5	BC063362 Xenopus t
17	561.8	45.4	177389	10	AC122453 Mus muscu
18	403.8	32.6	1996	5	BC055544 Danio rer
19	227.8	18.4	733	9	AF277190 Homo sapi

20	226.2	18.3	480	6	C0670930 Sequence
21	226.2	18.3	850	9	BC000383 Homo sapi
22	226.2	18.3	850	9	BC004432 Homo sapi
23	226.2	18.3	1470	6	C0725894 Sequence
24	226.2	18.3	1513	6	C0720118 Sequence
25	226.2	18.3	1600	6	AR116020 Sequence
26	226.2	18.3	1622	6	AX333010 Sequence
27	226.2	18.3	1622	9	HUMORF05 Homo sapien
28	226.2	18.3	1629	9	AK127822 Homo sapi
29	226.2	18.3	149679	9	AL158050 Human DNA
30	222.8	18.0	547	11	BV104914 MARC 1773
31	222.4	18.0	547	11	BV104914 MARC 1773
32	211.8	17.1	3244	10	BC046416 Mus muscu
33	211.8	17.1	58149	2	AC102560 Mouse DNA
34	211.8	17.1	185138	10	AL845172 Mouse DNA
35	201.2	16.3	475	6	C0459639 Sequence
36	200.5	16.2	187359	10	AC087183 Mus muscu
37	200.5	16.2	244215	2	AC073706 Mus muscu
38	195.8	15.8	201064	2	AC104750 Mus muscu
39	173.2	14.0	110000	2	AC074222.3 of
40	167.3	13.6	2224	3	DME243607 Drosophil
41	167.3	13.6	2552	3	DME243599 Drosophil
42	167.3	13.6	2553	6	C0585819 Sequence
43	167.3	13.6	2561	3	AY069478 Drosophil
44	147.4	11.9	820	5	BX935413 Gallus ga
45	147.4	11.9	843	5	CR353767 Gallus ga

ALIGNMENTS

RESULT 1
AX069004
LOCUS AX069004 1238 bp DNA linear PAT 25-JAN-2001
DEFINITION Sequence 12 from Patent WO0102569.
ACCESSION AX069004
VERSION AX069004.1 GI:12578824
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Hominidae; Homo.
REFERENCE
1
Tang, Y.T., Yue, H., Yang, J., Azimzai, Y., Baughn, M.R. and Lu, D.A.
Human immune response molecules
Patent: WO 0102569-A 12 11-JAN-2001;
JOURNAL Incyte Genomics, Inc. (US)
FEATURES
Location/Qualifiers
source
1..1238
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 100.0%; Score 1238; DB 6; Length 1238;
Best Local Similarity 100.0%; Pred. No. 1.3e-270; Mismatches 0; Indels 0; Gaps 0;
Matches 1238; Conservative 0;
QY 1 ACATCTTGGATGGGACTAGCAACCAAGGGACAAGATATTCAGAGTGTACTCTCAA 60
DB 1 ACATCTTGGATGGGACTAGCAACCAAGGGACAAGATATTCAGAGTGTACTCTCAA 60
QY 61 TCAGTACTCAAGCAAGTCCAGGAGCCTAGCGTTGCCCACTAGATCAACAATGGTAG 120
DB 61 TCAGTACTCAAGCAAGTCCAGGAGCCTAGCGTTGCCCACTAGATCAACAATGGTAG 120
QY 121 ACCCAGCGATCACTTGTGTTTTTCTTAAAAATGAAAGGTGAATGGAACAGACTAAAGACA 180
DB 121 ACCCAGCGATCACTTGTGTTTTTCTTAAAAATGAAAGGTGAATGGAACAGACTAAAGACA 180
QY 181 AACTGGAACAAGCCCAAAATGAAGTGGAGTTCCTGGAGTTTACGCTGTAGTACCAACAG 240
DB 181 AACTGGAACAAGCCCAAAATGAAGTGGAGTTCCTGGAGTTTACGCTGTAGTACCAACAG 240

QY	241	GGAAAAAGTTAATGCGAAGTGTGAAATGCTTATCCAGGAGAATCAAGAGCTTGGAGGC	300
Db	241	GGAAAAAGTTAATGCGAAGTGTGAAATGCTTATCCAGGAGAATCAAGAGCTTGGAGGC	300
QY	301	AGCTCTCCAGGAGCTATTGCACAACTTGAAGCAGAGTGGCTTTACAGAGAAATACA	360
Db	301	AGCTCTCCAGGAGCTATTGCACAACTTGAAGCAGAGTGGCTTTACAGAGAAATACA	360
QY	361	GTGAGGAGCTTAAAGAGCAGTCCAGATGAATGACTTCCATCATCCAGCTTGATCAAG	420
Db	361	GTGAGGAGCTTAAAGAGCAGTCCAGATGAATGACTTCCATCATCCAGCTTGATCAAG	420
QY	421	AAGTAGAGGGTATGCAGAGTACCATTCTAGTTCGACAGCAGCTGAAGAGACACGCC	480
Db	421	AAGTAGAGGGTATGCAGAGTACCATTCTAGTTCGACAGCAGCTGAAGAGACACGCC	480
QY	481	AGCAGTTGGCTCAGTACCAGCAGCAGTCTCAGGCTCTGCCCAAGTACACGAGGA	540
Db	481	AGCAGTTGGCTCAGTACCAGCAGCAGTCTCAGGCTCTGCCCAAGTACACGAGGA	540
QY	541	CTACAGCTTCTGAACCTGTAGAACAGTCCAGAGGCCACAAAGTAAAGACTGCAGTCTCTGA	600
Db	541	CTACAGCTTCTGAACCTGTAGAACAGTCCAGAGGCCACAAAGTAAAGACTGCAGTCTCTGA	600
QY	601	CAACCGGACCAAGTAATGGTAGTCTCTCCGCCAGAGAGCGTCTGGGTCTGGATTCA	660
Db	601	CAACCGGACCAAGTAATGGTAGTCTCTCCGCCAGAGAGCGTCTGGGTCTGGATTCA	660
QY	661	GGGAGGGCAACACAAACCAAGATCACTTTCCTCTCCAGGGAATGTAATAAGTCCT	720
Db	661	GGGAGGGCAACACAAACCAAGATCACTTTCCTCTCCAGGGAATGTAATAAGTCCT	720
QY	721	CCACAGCTCAGAGAGAGAACTGGCAGAGAGTAGTGTACGTAAATCAACTCAGTG	780
Db	721	CCACAGCTCAGAGAGAGAACTGGCAGAGAGTAGTGTACGTAAATCAACTCAGTG	780
QY	781	CGGGGTGAAGTGTAGACTCTCCACGGGAGTGAAGTCTCTCACACCAATCAA	840
Db	781	CGGGGTGAAGTGTAGACTCTCCACGGGAGTGAAGTCTCTCACACCAATCAA	840
QY	841	ATGACACAGACTCCAGTCAATGACCTCAAGAGGAGAAAGCAGTGGGAAAGGTAATC	900
Db	841	ATGACACAGACTCCAGTCAATGACCTCAAGAGGAGAAAGCAGTGGGAAAGGTAATC	900
QY	901	GAACTGTGGTTCGGCCAGCTCAGATGGCTTGGACTCAAGTGAATGTACAGGTT	960
Db	901	GAACTGTGGTTCGGCCAGCTCAGATGGCTTGGACTCAAGTGAATGTACAGGTT	960
QY	961	CAGTTTGTATAATTTTTCAGCAAAATTTTATACAGTGCATTTAATTTGGGAGAGGAT	1020
Db	961	CAGTTTGTATAATTTTTCAGCAAAATTTTATACAGTGCATTTAATTTGGGAGAGGAT	1020
QY	1021	ACTGTCAGAAAATTAATGCAATTTTGTGCAATTTGCTTTTGTGGGTGTACGTTT	1080
Db	1021	ACTGTCAGAAAATTAATGCAATTTTGTGCAATTTGCTTTTGTGGGTGTACGTTT	1080
QY	1081	TGGTTTTTTTGTGTTTTTTTTCATGATTTTTCCTTTCTTTTCTTTTATAGAGTGG	1140
Db	1081	TGGTTTTTTTGTGTTTTTTTTCATGATTTTTCCTTTCTTTTCTTTTATAGAGTGG	1140
QY	1141	TGTGTTTTTTTGTGTTTTTTTTCATGATTTTTCCTTTCTTTTCTTTTATAGTACTTT	1200
Db	1141	TGTGTTTTTTTGTGTTTTTTTTCATGATTTTTCCTTTCTTTTCTTTTATAGTACTTT	1200
QY	1201	GAAATGTCGAAAGGGCATTTTGTGGAGGTCAGTTA 1238	
Db	1201	GAAATGTCGAAAGGGCATTTTGTGGAGGTCAGTTA 1238	

RESULT 2
BC069192
LOCUS

DEFINITION

1. mRNA (CDNA clone MGC:78409 IMAGE:4653831), complete cds.

ACCESSION

BC069192

VERSION

BC069192.1

KEYWORDS

MGC.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 2108)

AUTHORS

Strausberg, R.D., Feingold, E.A., Grouse, L.H., Derge, J.G.,

Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,

Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,

Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,

Diachenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,

Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,

Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,

Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,

Abrahamson, R.D., Mullaly, S.J., Bosak, S.A., McGowan, P.J.,

McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,

Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,

Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,

Fahy, J., Helton, E., Kettman, M., Madan, A., Rodrigues, S.,

Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,

Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,

Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,

Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalls, D.E.,

Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length

human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932

2 (bases 1 to 2108)

Strausberg, R.

Direct Submission

Submitted (26-APR-2004) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgapbs-remail.nih.gov

Tissue procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Genome Sequence Centre,

BC Cancer Agency, Vancouver, BC, Canada

info@bcgsc.bc.ca

Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Butterfield,

Andy Chan, Steve S. Chand, William Chow, Alison Cloutier, Ruth

Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao,

Kim MacDonald, Anara Masson, Mike R. Mayo, Josh Moran, Ryan Morin,

Teika Olson, Diana Palmquist, Anca Petrescu, Anna Liisa Prabu,

Parvaneh Saeedi, JR Santos, Angelique Schreier, Ursula Skalska,

Duane Smalls, Jeff Stott, Miranda Tsai, George Yang, Jacquie

Schein, Asim Siddiqui, Rob Holt, Marco Marra.

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IPAL Plate: 54 Row: d Column: 11

This clone was selected for full length sequencing because it

passed the following selection criteria: matched mRNA gi: 23199975.

Location/Qualifiers

1. .2108

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="MGC:78409 IMAGE:4653831"

/tissue_type="placenta, chorioncarcinoma"

/lab_host="NIH MGC 21"

/lab_host="DH10B-R"

/note="Vector: pOTB7"

1. .2108

gene

	/gene="WTAP"	
	/note="synonyms: KIAA0105, MGC3925"	
	/db_xref="LocusID:9589"	
	/db_xref="MIM:605442"	
CDS	181..1371	
	/gene="WTAP"	
	/codon_start=1	
	/product="Wilms' tumour 1-associated protein, isoform 1"	
	/protein_id="AAH69192.1"	
	/db_xref="GI:46623320"	
	/db_xref="LocusID:9589"	
	/db_xref="MIM:605442"	
	/translation="MTNEBKLPPKVRILSETDFVKWMDLEILIRKWQYEAYVQAEGKY TDLNNDVTGLRESEKQLKOOCESARRNLVMRLATKEQMBCETTOIQYLKVQC PSVAQRSTNVDPALINFELFKMGGELEQFWDKLEQQNELSAKFVPDSDQTGKKLMX CRMLIQENQBLGRLOAEALAKYKBEEKLSQSDELNFIQLDBEVGM QSTLIVLQOOLKTROLAQYOQOQASAPSRTTBEPVSQTSATSKDCSURLNG PNSGSRRORTSGSHRGNTEITDDFPSPGPNKSINSSEERTGGSGSYQNLSA GYSDVSPTGSENSLTHQSNDTDSSHDPQBEKA VSGKNRTVGSRHVQNGLDSSVVNQ GSVL"	
ORIGIN		
	Query Match 94.9%; Score 1174.6; DB 9; Length 2108;	
	Best local Similarity 97.8%; Pred.No. 3.4e-256;	
	Matches 1212; Conservative 0; Mismatches 24; Indels 3; Gaps 2;	
Qy	1 ACATCCTTTGAATCGCACTAGCACCAAGGACAGAATATTCAAAGGTGTACTACTAAA 60	
Dd	401 ACATCCTTTGAATCGCACTAGCACCAAGGACAGAATATTCAAAGGTGTACTACTAAA 460	
Qy	61 TCCAGTAGCTCAAGCAAGTCAGCAGCCTACGCGTTGCCCACTTAGATCAACAATGGTAG 120	
Dd	461 TCCAGTAGCTCAAGCAAGTCAGCAGCAGCGGTGCCCCAACCTAGATCAACAATGGTAG 520	
Qy	121 ACCCAGCATCAACTCTGTTTTCTTAATAATGAAGGTGAACCTGGACAGACTAAAGACA 180	
Dd	521 ACCCAGCGATCAACTCTGTTTTCTTAATAATGAAGGTGAACCTGGAAACAAGCTAAAGACA 580	
Qy	181 AACTGGAAACAAGCCC AAAATGAACTGAGTGCCTGGAAATTTTACCCTCTGATAGCCAACAG 240	
Dd	581 AACTGGAAACAAG CCC AAATGAACTGAGTGCCTGGAAGTTTACGCCCTGATAGCCAACAG 640	
Qy	241 GGAAAAAGTTAATGGGAGTGTGGAATGCTTATCCAGGAGAACTCAAGAGCTTGGAGGC 300	
Dd	641 GGAAAAAGTTAATGGCGAAGTGTGCAATGCTTATCCAGGAGAACTCAAGAGCTTGGAGGC 700	
Qy	301 AGCTGTCCACGGAGCTATTGCCACAACTTGAAGCAGAGTGTGGCTTTACAGAAGAAATACA 360	
Dd	701 AGCTGTCCACGGAGCTATTGCCACAACTTGAAGCAGAGTGTGGCTTTACAGAAGAAATACA 760	
Qy	361 GTGAGGAGCTTTAAAGCAGCTCAGGATGAACCTGAATGATCTCATCCAGCTTGATGAAG 420	
Dd	761 GTGAGGAGCTTTAAAGCAGCTCAGGATGAACCTGAATGATCTCATCCAGCTTGATGAAG 820	
Qy	421 AAGTAGAGGGTATCGAGTAGTACCACTTAGTTCTTCGACAGCAGCTGAAGGAGACAGCC 480	
Dd	821 AAGTAGAGGGTATCGAGTAGTACCACTTAGTTCTTCGACAGCAGCTGAAGGAGACAGCC 880	
Qy	481 AGCAGTTGGCTCAGTACCAGCAGCAGCAGTCTCAGGCCTCTGCCCAAGTAGTACCAGCAGGA 540	
Dd	881 AGCAGTTGGCTCAGTACCAGCAGCAGCAGTCTCAGGCCTCTGCCCAAGTAGTACCAGCAGGA 940	
Qy	541 CTACAGCTTCTGAACCTGTAGAACAGTCAGAGGCCACAAGTAAGACTGCGAGTCTGTCGA 600	
Dd	941 CTACAGCTTCTGAACCTGTAGAACAGTCAGAGGCCACAAGTAAGACTGCGAGTCTGTCGA 1000	
Qy	601 CAAACGGACCAAGTAATGGTAGTCTCCTCCGCGCAGAGGACGTCTGGGTCTCGAATTTACA 660	
Dd	1001 CAAACGGACCAAGTAATGGTAGTCTCCTCCGCGCAGAGGACGTCTGGGTCTCGAATTTACA 1060	
Qy	661 GGGAGGGCAACAACACCGAAGTAGATTTCTTCTTCTTCTCCAGGGAACTGGTAATAGTCTT 720	
Dd	1061 GGGAGGGCAACAACACCGAAGTAGATTTCTTCTTCTTCTTCTCCAGGGAACTGGTAATAGTCTT 1120	

```

/protein_id="RAKS4764.1"
/db_xref="GI:14161495"
/translation="MINEPAPKKVRLSEDFKVMARDELILRWKQVEAVVQALEGKY
TDLNSNDVTGLRESEKLUKQOQESARENILVRLATKEQEMOECTTQIYLRKQVQ
PSVAOLRTMVDPAINFLLKMGLEQTKDLKLEQAQNELSAWKFDPDSOTGKKLMK
CRMLTQENQELGRQLSOGRIALQAEALAKLYSEELKSQDELNDPIIOLDEVEGM
OSTILVLOQQLRQQAQYQOQSOASAPSTRITASEPEOSEATSKDCSRLTNG
PSNGSSSRTSGSGFHEGNTTDDPSSPCNGKSSNSSEERTGSGSGVYNLSA
GVESVDSPTGSENSLTHQSDNTDSSHDPPQBEKAVSGKGNRTVSGRHVQNLDSRVVQ
GSVL"

```

ORIGIN

```

Query Match      93.7%; Score 1159.8; DB 9; Length 2134;
Best Local Similarity 97.3%; Pred. No. 7.7e-253;
Matches 1206; Conservative 0; Mismatches 17; Indels 16; Gaps 2;

QY 1 ACATCCITGGATGCGACTAGCAACCAAGGACAGAGATATTCAAGAGTGACTACTCAAA 60
DB 446 ACATCCITGGATGCGACTAGCAACCAAGGACAGAGATGCAAGAGTGACTACTCAAA 505
QY 61 TCCAGTACTCTCAAGCAAGTCCAGAGCGCTAGCGTTCGCCAATCTTAGATCAACATGTTAG 120
DB 506 TCCAGTACTCTCAAGCAAGTCCAGAGCGCGAGCGTTCGCCAATCTTAGATCAACATGTTAG 565
QY 121 ACCGAGCGATCAACTTTGTTTTTCTTAAATGAAAGGTGAACCTGGAACAGACTAAAGACA 180
DB 566 ACCGAGCGATCAACTTTGTTTTTCTTAAATGAAAGGTGAACCTGGAACAGACTAAAGACA 625
QY 181 AACTGGAAACAGCCCAAAATGAATGAGTGCTCGAAGTTTACGCTGATAGCCCAACAG 240
DB 626 AACTGGAAACAGCCCAAAATGAATGAGTGCTCGAAGTTTACGCTGATAGCCCAACAG 685
QY 241 GGAAAAAGTTAATGGCGAAGTGTGCAATGCTTATCCAGGAGAATCAAGAGCTTGGAGGC 300
DB 686 GGAAAAAGTTAATGGCGAAGTGTGCAATGCTTATCCAGGAGAATCAAGAGCTTGGAGGC 745
QY 301 AGCTGTCCCGGAGCGATTGACAACTGAGCACTGAGAGAGTGGCTTTACAGAAGAAATACA 360
DB 746 AGCTGTCCCGGAGCGATTGACAACTGAGCACTGAGAGAGTGGCTTTACAGAAGAAATACA 805
QY 361 GTGAGGAGCTTAAAGCAGTCCAGATGAATGATGACTTCACTCATCCAGCTTGATGAAG 420
DB 806 GTGAGGAGCTTAAAGCAGTCCAGATGAATGATGACTTCACTCATCCAGCTTGATGAAG 865
QY 421 AAGTAGAGGTTATCCAGATACCACTTCTAGTCTGAGAGAGCAGCTGAGAGAGACAGCC 480
DB 866 AAGTAGAGGTTATCCAGATACCACTTCTAGTCTGAGAGAGCAGCTGAGAGAGACAGCC 925
QY 481 AGCAGTTGGCTCAGTACAGCAGCAGCTCTCAGGCTCTGCCCAGCAAGTACCAGCAGGA 540
DB 926 AGCAGTTGGCTCAGTACAGCAGCAGCAGCTCTCAGGCTCTGCCCAGCAAGTACCAGCAGGA 985
QY 541 CTACAGCTTCTGAACCTGTGAGACAGTCCAGAGCCCAAGTAAGACTGACGTCGCTGA 600
DB 986 CTACAGCTTCTGAACCTGTGAGACAGTCCAGAGCCCAAGTAAGACTGACGTCGCTGA 1045
QY 601 CAAACGGACCAAGTAATGTTAGTCTCTCCCGCAGAGAGCTGCTGGTCTGGATTTACA 660
DB 1046 CAAACGGACCAAGTAATGTTAGTCTCTCCCGCAGAGAGAGCTGCTGGTCTGGATTTACA 1105
QY 661 GGGAGGGCAACACAAACCGAAGATGACTTCTTCTTCCAGGGAATGTTAAGTCTCT 720
DB 1106 GGGAGGGCAACACAAACCGAAGATGACTTCTTCTTCTCCAGGGAATGTTAAGTCTCT 1165
QY 721 CCAACAGCTCAGAGGAGAGACTGGCAGAGAGGTAGTGTACGTAAATCACTCACTCAGTG 780
DB 1166 CCAACAGCTCAGAGGAGAGACTGGCAGAGAGGTAGTGTACGTAAATCACTCACTCAGTG 1225
QY 781 CGGGGTAATGAAGTTAGACTCTCCACCGGCGAGTGAAGAACTCTCTCACACACCAATCAA 840
DB 1226 CGGGGTAATGAAGTTAGACTCTCTCCACCGGCGAGTGAAGAACTCTCTCACACACCAATCAA 1285
QY 841 ATGACACAGACTCCAGTCTGATGACCTTCAAGAGGAGAAAGCAGTGAGTGGAAGGTAATC 900

```

```

DB 1286 ATGACACAGACTCCAGTCAATGATCCCTCAAGAGGAGAAAGCAGTGAGTGGAAAGTAATC 1345
QY 901 GAACTGTGGGTTCCCGCCACCGTTTCAAGATGGCTTGGACTCAAGTGTAAATGTACAGGGTT 960
DB 1346 GAACTGTGGGTTCCCGCCACCGTTTCAAGATGGCTTGGACTCAAGTGTAAATGTACAGGGTT 1405
QY 961 CAGTTTGTAAATATTTTTCAGCAAAATTTTATACAGTGTCAATTTAATTTGGGAGAGAT 1020
DB 1406 CAGTTTGTAAATATTTTTCAGCAAAATTTTATACAGTGTCAATTTAATTTGGGAGAGAT 1465
QY 1021 ACTGTCCAGAAAATTAATGCATACCTTTTGTCAAAATTTGCCCTTTTGTGGGTGTACGTTT 1080
DB 1466 ACTGTCCAGAAAATTAATGCATACCTTTTGTCAAAATTTGCCCTTTTGTGGGTGTACGTTT 1525
QY 1081 TGGTTTTTTTTTGTGTTTTTTTTTCTTGGAAATTTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1140
DB 1526 TGGTTTTTTTTTGTGTTTTTTTTTCTTGGAAATTTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1576
QY 1141 TTGTTGTTTTTTTGTCTTCAATACCTTCTGCGCTTGGGAATTTGTAACAGTAAATACCTTT 1200
DB 1577 -----TTTTTTTGTCTTCAATACCTTCTGCGCTTGGGAATTTGTAACAGTAAATACCTTT 1630
QY 1201 GAATGTTC-AAAAGGCGATTTTGTGGAGGTCAGTTA 1238
DB 1631 GAATGTTC-AAAAGGCGATTTTGTGGAGGTCAGTTA 1669

RESULT 4
MMU276707
LOCUS      Mus musculus partial mRNA for WTAP protein.
DEFINITION
ACCESSION  AJ276707
VERSION     AJ276707.1  GI:11322454
KEYWORDS    Wtap gene; WTAP protein.
SOURCE      Mus musculus
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE   1  Little,N.A.
            2  Thesis (1999) Edinburgh University
            3  Little,N.A., Hastie,N.D. and Davies,R.C.
            4  Identification of WTAP, a novel Wilms' tumour 1-associating protein
            5  Hum. Mol. Genet. 9 (15), 2231-2239 (2000)
            6  MEDLINE 20458888
            7  PUBMED 11001926
            8  REFERENCE 3 (bases 1 to 1802)
            9  AUTHORS Hastie,N.D.
           10  TITLE Direct Submission
           11  JOURNAL Submitted (22-MAR-2000) Hastie N.D., Developmental Biology, MRC
           12  Human Genetics Unit, Western General Hospital, Crewe Road,
           13  Edinburgh, EH4 2XU, UNITED KINGDOM
           14  FEATURES
           15  source
           16  1..1802
           17  /organism="Mus musculus"
           18  /mol_type="mRNA"
           19  /db_xref="taxon:10090"
           20  /clone_lib="Clontech mouse 11 day embryo 5' STRETCH PLUS
           21  cDNA library"
           22  1..1802
           23  /gene="wtap"
           24  /locus_tag="1140"
           25  /gene="wtap"
           26  /codon_start=1
           27  /product="WTAP protein"
           28  /protein_id="CAC16790.1"
           29  /db_xref="GI:11322455"
           30  /db_xref="GOA:Q8ER69"
           31  /db_xref="TrEMBL:Q8ER69"
           32  /translation="FKVMARDELILRWKQVEAVVQALEGKYTDLNSNDVTGLRESEBEK
           33  LKQOQOQESARENILVRLATKEQEMOECTTQIYLRKQVQPSVAOLRTMVDPAINFLLKMGLEQTKDLKLEQAQNELSAWKFDPDSOTGKKLMK
           34  CRMLTQENQELGRQLSOGRIALQAEALAKLYSEELKSQDELNDPIIOLDEVEGMOSTILVLOQQLRQQAQYQOQSOASAPSTRITASEPEOSEATSKDCSRLTNG
           35  PSNGSSSRTSGSGFHEGNTTDDPSSPCNGKSSNSSEERTGSGSGVYNLSAGVESVDSPTGSENSLTHQSDNTDSSHDPPQBEKAVSGKGNRTVSGRHVQNLDSRVVQ
           36  GSVL"

```

FFLMKGELEOTKDKLEQONELSAWKFTPDSTQGTGKLMKCRMLIOENOLGROLSSQ
GRIQAEALALOKKYSEELKSSODELNDPILCLDEVEGMSSTILVLOQQLKREFO
LAQYQOQSOASAPSTTSSEPVDAEVTSCORLANGSPNSGSSRSGSGGFH
REGSTPDDPSSNGNXSNSEERTGSGSSYIPLSAGTSEYSDPTSGNSLTH
HSNDTDSHDPQBEKAVSGKGNRTVGRSHVQNGLDSSVNVQAVL

ORIGIN

Query Match 79.2%; Score 980; DB 10; Length 1802;
Best Local Similarity 89.2%; Pred. No. 5.5e-212; Indels 38; Gaps 3;
Matches 1105; Conservative 0; Mismatches 96;

QY 1 ACATCCCTTGAATGCGACTAGCAACCAAGGAGCAAGATATTCAGAGGTGCTACTCAAA 60
DB 170 ACATCTTGTATGTCGCGCTAGCAACCAAGGAGGAGATGCAAGGTGCAACACTCAA 229
QY 61 TCCAGTACCTCAAGCAAGTCCAGAGCTAGCGTTCGCCAATCTAGATCAACAATGGTAG 120
DB 230 TCCAGTACCTCAAGCAAGTCCAGAGCTAGCGTTCGCCAATCTAGATCAACAATGGTAG 289
QY 121 ACCAGAGGATCAACTGTCTTTCTCTAAATGAAAGGTGAAGTGGAAACAGACTAAAGACA 180
DB 290 ACCAGAGGATCAACTGTCTTTCTCTAAATGAAAGGTGAAGTGGAAACAGACTAAAGACA 349
QY 181 AACTGGAAACAGCCCAAAATGAATGAGTGTCTGGAAGTTTACGCTGTATAGCCAAACAG 240
DB 350 AACTGGAAACAGCCCAAAATGAATGAGTGTCTGGAAGTTTACGCTGTATAGCCAAACAG 409
QY 241 GGAAGAGTTAATGCGAGGTGTCGAATGCTTATCCAGGAGAAATCAAGAGCTTGGAGGC 300
DB 410 GCAAGAGCTAATGGCAAGTGTGCAATGCTTATCCAGGAGAACCAAGAGCTTGGAGGC 459
QY 301 AGCTGTCCAGGAGCTATTGACACAACTTGAAGCAGAGTTGGCTTTACAGAAATACA 360
DB 470 AGCTGTCCAGGAGCTATTGACACAACTTGAAGCAGAGTTGGCTTTACAGAAATACA 529
QY 361 GTGAGGAGCTTAAAGCAGTCCAGATGAATGATCTCATCATCCAGCTGTATGAAG 420
DB 530 GTGAGGAGCTTAAAGCAGTCCAGATGAATGATCTCATCATCCAGCTGTATGAAG 589
QY 421 AAGTAGAGGATGATGAGAGTACCAATCTTAGTTCTGCGAGCAGCTGAAGGAGACGCC 480
DB 590 AAGTAGAGGATGATGAGAGTACCAATCTTAGTTCTGCGAGCAGCTGAAGGAGACGCC 649
QY 481 AGCAGTTGGCTCAGTACAGCAGCAGAGTCTCAGGCTCTGCCGCCAGTACCGAGCA 540
DB 650 AGCAGTTGGCTCAGTACAGCAGCAGAGTCTCAGGCTCTGCCGCCAGTACCGAGCA 709
QY 541 CTACAGCTTCTGAACCTGTAGAACAGTCCAGAGGCCAACAGTAAAGACTGCGTCTGA 600
DB 710 CTACATCTTCTGAACCTGTAGAACAGTCCAGAGGCCAACAGTAAAGACTGCGTCTGA 769
QY 601 CAACGGACCAAGTAAATGGTAGCTCTCCCGCAGAGAGAGCTCTGGGTCTGGATTACA 660
DB 770 CAATGGACCAAGTAAATGGTAGCTCTCCCGCAGAGAGAGCTCTGGGTCTGGATTACA 829
QY 661 GGGAGGCAACACACCCAGAGTACTTCTCTCTCCAGGGAATGGTAATAGCTCT 720
DB 830 GGGAGGCAACACACCCAGAGTACTTCTCTCTCTCCAGGGAATGGTAATAGCTCT 889
QY 721 CCAACAGCTCAGAGGAGAACTGGCAGAGGAGGTAGTGTTCGTAATCAATCACTAGT 780
DB 890 CCAACAGCTCAGAGGAGAACTGGCAGAGGAGGTAGTGTTCGTAATCAATCACTAGT 949
QY 781 CGGGGTATGAAGTGTAGCTCTCCAGGGGAGTGAATCTCTCTACACCACTCAAA 840
DB 950 CGGGGTATGAAGTGTAGCTCTCCAGGGGAGTGAATCTCTCTACACCTCACTCAA 1009
QY 841 ATGACACAGCTCCAGTCAATGACCTCAAGAGGAGAGAGAGTGTAGTGGGAAAGTATC 900
DB 1010 ATGACACAGCTCCAGTCAATGACCTCAAGAGGAGAGAGAGTGTAGTGGGAAAGTATC 1069
QY 901 GAACTGTGGGTTCCCGCCACGTTTCAGATGGCTTGGACTCAAGTGTAAATGTACAGGTT 960

DB 1070 GAACTGTGGTTCGCCCATGTTCCAGATGCTTGGACTCAAGTGAATGTACAGGGTG 1129
QY 961 CAGTTTGTATATATTTTTCAGCAAAATTTTATACAGTGTCAATTTAATTTGGGAGAGAT 1020
DB 1130 CAGTTTGTATATATTTTTCAGCAAAATTTTATACAGTGTCAATTTAATTTGGGAAAGGA 1189
QY 1021 ACTGTCCAGAAAATTAATGCACTATTTGTACAAATTTGCCCTTTTGTGGGTGTACGTTT 1080
DB 1190 TACTGTCAAAAATTAATGCACTATTTTGTACAAATTTACCTTTTGTGGGTGT--GTTT 1247
QY 1081 TGGTTTTTTTGT 1275
DB 1248 TTGTTTGT 1275
QY 1141 TTGTTGT 1275
DB 1276 ---TTTTTAAATGCTTCAATACCTCTGCCACTTTGAAAATTTGTAACAGTTAAATFACTTT 1332
QY 1201 GAATGTTGC-AAAAGGGCATTTTGTGGAGGTCAGTTA 1238
DB 1333 GAACATTTGCTAAAGGACATTTTGTAGGGTCAAGTTA 1371

RESULT 5

CQ845798 2291 bp DNA linear PAT 02-AUG-2004
Sequence 4445 from Patent EP1440981.
DEFINITION
CQ845798
ACCESSION
CQ845798.1 GI:50895397
VERSION
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
AUTHORS
Isogai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S.,
Yamamoto, J., Isono, Y., Nagai, K. and Irie, R.
TITLE
Full-length human cdna
JOURNAL
Patent: EP 1440981-A 4445 28-JUL-2004;
Research Association for Biotechnology (JRP)
FEATURES
Location/Qualifiers
source
1..2291
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 77.0%; Score 953.8; DB 6; Length 2291;
Best Local Similarity 97.9%; Pred. No. 5e-206;
Matches 988; Conservative 0; Mismatches 17; Indels 4; Gaps 2;

QY 231 AGCCAAACAGGGAAAAGTTAATGGGAAAGTGTGGAATGCTTATCCAGGAGAAATCAAGAG 290
DB 841 AGCCAAACAGGGAAAAGTTAATGGGAAAGTGTGGAATGCTTATCCAGGAGAAATCAAGAG 900
QY 291 CTTTGAAGCAGAGCTGTCCAGGAGACGTATTGCAACAACTTGAAGCAGAGTTGGCTTTACAG 350
DB 901 CTTTGAAGCAGAGCTGTCCAGGAGACGTATTGCAACAACTTGAAGCAGAGTTGGCTTTACAG 960
QY 351 AAGAAATACAGTACAGAGCTTAAAGCAGTCCAGGATGAATGACTCATCATCCAG 410
DB 961 AAGAAATACAGTACAGAGCTTAAAGCAGTCCAGGATGAATGACTCATCATCCAG 1020
QY 411 CTTGATGAAGAAGTAGAGGATGTCAGAGTACCAATCTAGTTCTGTGAGCAGCAGCTGAAG 470
DB 1021 CTTGATGAAGAAGTAGAGGATGTCAGAGTACCAATCTAGTTCTGTGAGCAGCAGCTGAAG 1080
QY 471 GAGACACGCCAGAGTTGGCTCAGTACAGCAGCAGCAGCTCTCAGGCTCTGCCCAAGT 530
DB 1081 GAGACACGCCAGAGTTGGCTCAGTACAGCAGCAGCAGCTCTCAGGCTCTGCCCAAGT 1140
QY 531 ACCAGACAGGACTACAGCTTCTGAACCTGTAGAACAGTCCAGAGCCCAAGTAAAGACTGC 590

Db	1501	AAAGGTAATCGAACTGTGGGTTCCGCCACGTTCAAGATGGCTTGGACTCAAGTGAAT	1566
Qy	951	GTACAGGGTTCAGTTTGTATAATTTTTCAGCAAAATTTTATACAGTGTCATTAATTT	1010
Db	1561	GTACAGGGTTCAGTTTGTATAATTTTTCAGCAAAATTTTATACAGTGTCATTAATTT	1620
Qy	1011	GGGAGAGGATACGTCCAGAAAATTAATGATACTTTTGTCAAAATTTGCCTTTTGTGG	1070
Db	1621	GGGAGAGGATACGTCCAGAAAATTAATGATACTTTTGTCAAAATTTGCCTTTTGTGG	1680
Qy	1071	GTGTACGTTTGGTTTTTTTTTGTGTTTTTTTTTCTTTTGGATTTTTTTTCTTTCTTTTT	1130
Db	1681	GTGTACGTTTGGTTTTTTTTTGTGTTTTTTTTTCTTTTGGATTTTTTTTCTTTCTTTTT	1740
Qy	1131	TAAGAGGTGTTGTGTTTTTTTTTGCTTCAATCTCTTCGCGCTTGGGAATTTGTAACAGT	1190
Db	1741	T---TTTTTTTTTTTTTTTTTTTGTCTTCAATACTCTTCGCGCTTGGAAATTTGTAACAGT	1797
Qy	1191	TAATTACTTGAATGTTGC-AAAAGGCAATTTTGTGGAGGGTCAGTTA	1238
Db	1798	TAATTACTTTGAATGTTGCTAAAAGGACATTTTGTAGGCTCAAGTTA	1846

RESULT 7			
HSA276706			
LOCUS	HSA276706	1167 bp	linear
DEFINITION	Homo sapiens partial mRNA for WRAP protein.		
			PRI 26-SEP-2000

ORGANISM	<p>1. Homo sapiens</p> <p>2. Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.</p>
REFERENCE	<p>1. Little,N.A.</p> <p>2. Thesis (1999) Edinburgh University</p>
AUTHORS	<p>1. Little,N.A., Hastie,N.D. and Davies,R.C.</p>
REFERENCE	<p>2. Identification of WRAP, a novel Wilms' tumour 1-associating protein</p>
TITLE	<p>Hum. Mol. Genet. 9 (15), 2231-2239 (2000)</p>
JOURNAL	<p>Hum. Mol. Genet. 9 (15), 2231-2239 (2000)</p>

ORIGIN

RESULT 8
AL135914
LOCUS

Accession	Size	Library	Accession
AL135914	165802 bp	DNA	PRI 20-JUL-2001
LOCUS			linear

Query Match	76.2%	Score 943.8	DB 9	Length 1167
Best Local Similarity	99.3%	Pred. No. 9.2e-204		
Matches 948	Conservative 0	Mismatches 7	Indels 0	Gaps 0
QY 1	ACATCCTTTGGAAATGCGACTAGCAACCAAGGGAACAAGATATTCAAGAGTGTACTACTCAA	60		
DB 213	ACATCCTTTGGTAATGCGACTAGCAACCAAGGAACAAGAGATGCAAGAGTGTACTACTCAA	272		
QY 61	TCCAGTACTCAAGCAAGTCCAGCAGCCTAGCGTTGCCCACTTAGATCAACAATGGTAG	120		
DB 273	TCCAGTACTCAAGCAAGTCCAGCAGCAGCGTTGCCCACTAGATCAACAATGGTAG	332		
QY 121	ACCCAGCGATCAACTTGTCTTAAATAAGAAAGTGAACTGGAACAGACATAAGACA	180		
DB 333	ACCCAGCGATCAACTTGTCTTAAATAAGAAAGTGAACTGGAACAGACATAAGACA	392		
QY 181	AACTGGACACAGCCCAAAATGAACCTGAGTGCCTGGAAGTTTACGCTGTAGTACCAAAACAG	240		
DB 393	AACTGGACACAGCCCAAAATGAACCTGAGTGCCTGGAAGTTTACGCTGTAGTACCAAAACAG	452		
QY 241	GGAAAAAGTTAATGGCGAAGTCTCGAATGCTTATCCAGGAGAAATCAAGAGCTTGAAGGC	300		
DB 453	GGAAAAAGTTAATGGCGAAGTCTCGAATGCTTATCCAGGAGAAATCAAGAGCTTGAAGGC	512		
QY 301	AGCTGTCCAGGAGCGTATTGCACAACTTGAGCAGAGTTGCTTTACAGAGAAATACA	360		
DB 513	AGCTGTCCAGGAGCGTATTGCACAACTTGAGCAGAGTTGCTTTACAGAGAAATACA	572		
QY 361	GTGAGGAGCTTAAAGACAGTCAAGATGAATCAATGACTTCATCATCCAGCTTGATGAAG	420		
DB 573	GTGAGGAGCTTAAAGACAGTCAAGATGAATCAATGACTTCATCATCCAGCTTGATGAAG	632		
QY 421	AAGTAGAGGATGACAGATGACATTCCTAGTTCTGCAGCAGCAGCTGAAGGACACGCC	480		
DB 633	AAGTAGAGGATGACAGATGACATTCCTAGTTCTGCAGCAGCAGCTGAAGGACACGCC	692		
QY 481	AGCAGTTGGCTCAGTACCAGCAGCAGCTCTCAGGCTCTGCCCAAGTACACAGCAGGA	540		
DB 693	AGCAGTTGGCTCAGTACCAGCAGCAGCTCTCAGGCTCTGCCCAAGTACACAGCAGGA	752		
QY 541	CTACAGCTTCTGAAACCTGTAGAACAGTCAAGGCCCAAGTAAAGACTCGAGTCTGTGA	600		
DB 753	CTACAGCTTCTGAAACCTGTAGAACAGTCAAGGCCCAAGTAAAGACTCGAGTCTGTGA	812		
QY 601	CAAAACGACCAAGTAAATGTAGTCTCTCCCGCCAGAGGACGCTCGGCTCTGGATTTTACA	660		
DB 813	CAAAACGACCAAGTAAATGTAGTCTCTCCCGCCAGAGGACGCTCGGCTCTGGATTTTACA	872		
QY 661	GGAGGGCAACAAACCGAAGATGATTTCTTCTTCTCCAGGGAATGGTAATAGTCTCT	720		
DB 873	GGAGGGCAACAAACCGAAGATGATTTCTTCTTCTTCTCCAGGGAATGGTAATAGTCTCT	932		
QY 721	CCAAACAGCTCAGAGGAGAGAACTGGCAGAGGAGGTAGTGGTACGTAATACTCAGTG	780		
DB 933	CCAAACAGCTCAGAGGAGAGAACTGGCAGAGGAGGTAGTGGTACGTAATACTCAGTG	992		
QY 781	CGGGGTATGAAAGTGTAGACTCTCCACGGCAGTGAATACTCTCTCACACCAATCAA	840		
DB 993	CGGGGTATGAAAGTGTAGACTCTCCACGGCAGTGAATACTCTCTCACACCAATCAA	1052		
QY 841	ATGACACAGACTCCAGTCAATGACCTCCAGAGGAGAAAGCAGTGAAGTGGGAAAGGTAATC	900		
DB 1053	ATGACACAGACTCCAGTCAATGACCTCCAGAGGAGAAAGCAGTGAAGTGGGAAAGGTAATC	1112		
QY 901	GAACTGTGGGTTCCCGCCAGCTTCAGAAATGGCTTGGACTCAAGTGAATGTACA	955		
DB 1113	GAACTGTGGGTTCCCGCCAGCTTCAGAAATGGCTTGGACTCAAGTGAATGTACA	1167		

gene
CDS

FEATURES
source

REFERENCE	REFERENCE
AUTHORS	AUTHORS
JOURNAL	JOURNAL
REFERENCE	REFERENCE
AUTHORS	AUTHORS
TITLE	TITLE
JOURNAL	JOURNAL
MEDLINE	MEDLINE
PUBMED	PUBMED
REFERENCE	REFERENCE
AUTHORS	AUTHORS
TITLE	TITLE
JOURNAL	JOURNAL

RESULT 7
HSA276706
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

QY 1
Db 1
QY 1
Db 1
QY 1
Db 1

db	1
Qy	1
db	1
Qy	1
db	1

DEFINITION Human DNA sequence from clone RPI-56L9 on chromosome 6, complete sequence.

ACCESSION AL135914

VERSION AL135914.25 GI:15020718

KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 165802)

AUTHORS Williams, S.

TITLE Direct Submission

JOURNAL Submitted (20-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

COMMENT requests: clonerequest@sanger.ac.uk
On Jul 25, 2001 this sequence version replaced gi:14970777.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6>
RPI-56L9 is from the library RPCI-1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
VECTOR: pCYPAC2
IMPORTANT: This sequence is not the entire insert of clone RPI-56L9. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.
The true left end of clone RPI-56L9 is at 1 in this sequence. The true left end of clone RPI-195P10 is at 163803 in this sequence. The true right end of clone RPI1-280I21 is at 123148 in this sequence.

FEATURES

Source

1..165802
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RPI-56L9"
/clone_lib="RPCI-1"
399..564
/note="MIR repeat: matches 92. .262 of consensus"
714..1033
/note="AluY repeat: matches 1. .303 of consensus"
1293..1732
/note="L1PA16 repeat: matches 5711. .6157 of consensus"
1755..2035
/note="AluX repeat: matches 23. .306 of consensus"
2036..2143
/note="54 copies 2 mer ga 72% conserved"
2146..2466
/note="AluX repeat: matches 1. .310 of consensus"
2563..2660
/note="MLT1J repeat: matches 413. .513 of consensus"
4175..4463
/note="AluSg repeat: matches 1. .290 of consensus"

repeat_region
5359..5641
/note="AluX repeat: matches 3. .300 of consensus"
6418..6499
/note="L1MEC repeat: matches 2345. .2063 of consensus"
6500..6807
/note="AluJb repeat: matches 1. .312 of consensus"
6808..6934
/note="L1MEC repeat: matches 2062. .2195 of consensus"
7704..7992
/note="AluSg repeat: matches 18. .306 of consensus"
7994..8204
/note="L1ME1 repeat: matches 5236. .5402 of consensus"
8205..8418
/note="AluJb repeat: matches 87. .303 of consensus"
8520..8623
/note="L1ME1 repeat: matches 5629. .5729 of consensus"
8673..8948
/note="AluX repeat: matches 1. .293 of consensus"
9337..9639
/note="AluX repeat: matches 1. .303 of consensus"
9752..9977
/note="L1ME1 repeat: matches 5734. .5963 of consensus"
9981..10097
/note="LTR29 repeat: matches 515. .619 of consensus"
10040..10107
/note="MER39b repeat: matches 482. .552 of consensus"
10360..10560
/note="LTR29 repeat: matches 25. .215 of consensus"
10620..10811
/note="L1ME1 repeat: matches 5959. .6154 of consensus"
10929..11002
/note="37 copies 2 mer at 71% conserved"
11195..11504
/note="AluJo repeat: matches 6. .298 of consensus"
12430..13032
/note="LTR9 repeat: matches 12. .625 of consensus"
16460..16765
/note="AluJb repeat: matches 1. .305 of consensus"
16771..16824
/note="27 copies 2 mer ct 81% conserved"
17164..17468
/note="AluX repeat: matches 1. .304 of consensus"
17883..17894
/note="16 copies 2 mer ac 100% conserved"
18164..18237
/note="L2 repeat: matches 2670. .2749 of consensus"
18238..18548
/note="AluJo repeat: matches 1. .312 of consensus"
18549..18675
/note="L2 repeat: matches 2565. .2670 of consensus"
19011..19218
/note="AluSg/x repeat: matches 85. .290 of consensus"
19564..19902
/note="L2 repeat: matches 2398. .2750 of consensus"
20029..20158
/note="FLAM C repeat: matches 1. .107 of consensus"
20192..20422
/note="MIR repeat: matches 13. .242 of consensus"
21398..21587
/note="L2 repeat: matches 2482. .2727 of consensus"
22166..22927
/note="MER45B repeat: matches 250. .1039 of consensus"
22928..23247
/note="AluX repeat: matches 1. .311 of consensus"
23248..23502
/note="MER45B repeat: matches 2. .250 of consensus"
23565..24972
/note="704 copies 2 mer at 81% conserved"
24081..24105
/note="Tandem repeat. Single clone region. Assembly confirmed by restriction digest data"
24975..25042
/note="34 copies 2 mer ta 73% conserved"


```
repeat_region 25044..25357
/note="L1MC4 repeat: matches 6605..6822 of consensus"
repeat_region 25371..26043
/note="L1MC4 repeat: matches 7167..7977 of consensus"
repeat_region 26147..26273
/note="FLAM_A repeat: matches 19..142 of consensus"
repeat_region 26998..27598
/note="L1PBA repeat: matches -1271..-531 of consensus"
repeat_region 27661..30938
/note="L1PBI repeat: matches -275..2897 of consensus"
repeat_region 30939..31236
/note="AluSg repeat: matches 1..296 of consensus"
repeat_region 31237..312134
/note="L1PBI repeat: matches 2897..3766 of consensus"
repeat_region 32135..32930
/note="SVA repeat: matches 1..791 of consensus"
repeat_region 32931..33685
/note="L1PBI repeat: matches 3766..5563 of consensus"
repeat_region 33694..34341
/note="L1PBI repeat: matches 17..670 of consensus"
repeat_region 34342..34477
/note="L1PBI repeat: matches 5559..5681 of consensus"
repeat_region 34478..34782
/note="AluSg repeat: matches 1..306 of consensus"
repeat_region 34783..34889
/note="L1PBI repeat: matches 5681..5791 of consensus"
repeat_region 34892..34927
/note="Alu repeat: matches 1..36 of consensus"
repeat_region 34931..35025
/note="L1PBI repeat: matches 5687..5791 of consensus"
repeat_region 35062..35339
/note="139 copies 2 mer at 63% conserved"
repeat_region 35340..35699
/note="L1PBI repeat: matches 5781..6152 of consensus"
repeat_region 35762..35807
/note="23 copies 2 mer ct 91% conserved"
repeat_region 35850..35973
/note="L1PBI repeat: matches 243..374 of consensus"
repeat_region 36221..36687
/note="L2 repeat: matches 2165..2726 of consensus"
repeat_region 36812..36888
/note="L2 repeat: matches 1910..1982 of consensus"
repeat_region 36889..37175
/note="AluSp repeat: matches 1..296 of consensus"
repeat_region 37176..37379
/note="L2 repeat: matches 1689..1910 of consensus"
repeat_region 38041..38195
/note="AluDb repeat: matches 140..284 of consensus"
repeat_region 38304..38462
/note="AluJo repeat: matches 1..134 of consensus"
repeat_region 38859..38896
/note="MERSB repeat: matches 4..35 of consensus"
repeat_region 38897..39182
Query Match 64.4%; Score 797.8; DB 9; Length 165802;
Beat Local Similarity 97.5%; Pred. NO. 1.6e-170;
Matches 832; Conservative 0; Mismatches 17; Indels 4; Gaps 2;

QY 367 GAACGTGAATGACTTCATCATCATCAGCTTGATGAAGAGTAGAGGGTATGCAGAGTACCATT 446
DB 126710 GAACTGAATGACTTCATCATCATCAGCTTGATGAAGAGTAGAGGGTATGCAGAGTACCATT 126769

QY 447 CTAGTTCTGCAGCAGAGCTGAAGAGACAGCCAGCAGTGGCTCAGTACACAGCAG 506
DB 126770 CTAGTTCTGCAGCAGAGCTGAAGAGACAGCCAGCAGTGGCTCAGTACACAGCAG 126829

QY 507 CAGTCTCAGGCTCTGCCCAAGTACCAGCAGTCTGAGACCTGTGAACAG 566
DB 126830 CAGTCTCAGGCTCTGCCCAAGTACCAGCAGTCTGAGACCTGTGAACAG 126889

QY 567 TCAGAGGCCACCAAGTAAGACTGCAGTCTGTGACAAACCGACCAAGTAATGGTAGTCC 626
DB 126890 TCAGAGGCCACCAAGTAAGACTGCAGTCTGTGACAAACCGACCAAGTAATGGTAGTCC 126949
```

```
QY 627 TCCGCCAGAGGAGCGTCTGGGTCTGGATTTCACAGGAGGGCAACACACCGAGATGAC 686
DB 126950 TCCGCCAGAGGAGCGTCTGGGTCTGGATTTCACAGGAGGGCAACACACCGAGATGAC 127009

QY 687 TTTCCTCTCTCCAGAGGAATGTAATAAGTCTCCCAACAGCTCAGAGGAGAGAACTGGC 746
DB 127010 TTTCCTCTCTCCAGAGGAATGTAATAAGTCTCCCAACAGCTCAGAGGAGAGAACTGGC 127069

QY 747 AGAGGAGTAGTGGTTACGTAATCACTCAGTCCGGGGTATGAAGTGTAGACTCTCC 806
DB 127070 AGAGGAGTAGTGGTTACGTAATCACTCAGTCCGGGGTATGAAGTGTAGACTCTCC 127129

QY 807 AGGSCAGTGAATACTCTCTCACACCAATCAATGACACAGACTCCAGTCAATGACCCCT 866
DB 127130 AGGSCAGTGAATACTCTCTCACACCAATCAATGACACAGACTCCAGTCAATGACCCCT 127189

QY 867 CAAGAGGAGAAAGCAGTCAAGTGGGAAAGGTAAATCGAACTGGGTCCGCCACGTTGAG 926
DB 127190 CAAGAGGAGAAAGCAGTCAAGTGGGAAAGGTAAATCGAACTGGGTCCGCCACGTTGAG 127249

QY 927 AATGCTTGGACTCAAGTGAATGACAGGTTTCAGTTTGTATATATTTTTCAGCAA 986
DB 127250 AATGCTTGGACTCAAGTGAATGACAGGTTTCAGTTTGTATATATTTTTCAGCAA 127309

QY 987 TTTTATACAGTGTCAATTAATTTGGGAGAGGATACTGTCCAGAAAAATTAATGCATCTT 1046
DB 127310 TTTTATACAGTGTCAATTAATTTGGGAGAGGATACTGTCCAGAAAAATTAATGCATCTT 127369

QY 1047 TTGTCACAATTTGCCCTTTTGGGGTGTACGTTTGGTTTGTATATATTTTGTCT 1106
DB 127370 TTGTCACAATTTGCCCTTTTGGGGTGTACGTTTGGTTTGTATATATTTTGTCT 127429

QY 1107 TTGGATTTTTTTCTTTCTTTTAAAGAGGTGGTGTGTGTGTGTGTGTGTGTGTGTGT 1166
DB 127430 TTGGATTTTTTTCTTTCTTTTAAAGAGGTGGTGTGTGTGTGTGTGTGTGTGTGTGT 127486

QY 1167 TGGCGCTTGGGAATGTAAACAGTTAAATTAATTAATTAATTAATTAATTAATTAATTA 1225
DB 127487 TGGCGCTTGGGAATGTAAACAGTTAAATTAATTAATTAATTAATTAATTAATTAATTA 127546

QY 1226 GAGGTCACAGTTA 1238
DB 127547 TAGGGTCAAGTTA 127559

RESULT 9
AC127172 203077 bp DNA linear ROD 16-SEP-2003
LOCUS Mus musculus clone rp23-387d2 map 17 strain C57BL/6J, complete
DEFINITION sequence.
AC127172
VERSION AC127172.13 GI:30023881
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 203077)
AUTHORS Wiley,G., Yang,L., Prescott,A., Qin,B., Brathwaite,M., Waeltz,P.,
Nagazawa,R. and Roe,B.A.
Nagazawa,R. and Roe,B.A.
Mus musculus BAC Clone rp23-387d2
TITLE Unpublished
JOURNAL 2 (bases 1 to 203077)
REFERENCE 2 (bases 1 to 203077)
AUTHORS Wiley,G., Yang,L., Prescott,A., Qin,B. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (14-JUL-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 3 (bases 1 to 203077)
AUTHORS Wiley,G., Yang,L., Prescott,A., Qin,B. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (26-NOV-2002) Department Of Chemistry And Biochemistry,
```

The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
 4 (bases 1 to 203077)
 Wileg,G., Yang,L., Prescott,A., Qin,B. and Roe,B.A.
 Direct Submission
 Submitted (06-DEC-2002) Department Of Chemistry And Biochemistry,
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
 OK 73019, USA
 5 (bases 1 to 203077)
 Wileg,G., Yang,L., Prescott,A., Qin,B., Brathwaite,M., Waelz,P.,
 Nagaraja,R. and Roe,B.A.
 Direct Submission
 Submitted (21-MAR-2003) Department Of Chemistry And Biochemistry,
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
 OK 73019, USA
 6 (bases 1 to 203077)
 Wileg,G., Yang,L., Prescott,A., Qin,B., Brathwaite,M., Waelz,P.,
 Nagaraja,R. and Roe,B.A.
 Direct Submission
 Submitted (18-APR-2003) Department Of Chemistry And Biochemistry,
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
 OK 73019, USA
 7 (bases 1 to 203077)
 Wileg,G., Yang,L., Prescott,A., Qin,B., Brathwaite,M., Waelz,P.,
 Nagaraja,R. and Roe,B.A.
 Direct Submission
 Submitted (16-SEP-2003) Department Of Chemistry And Biochemistry,
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
 OK 73019, USA
 On Apr 18, 2003 this sequence version replaced gi:26080527.
 Center: Department Of Chemistry And Biochemistry
 The University Of Oklahoma
 Center code:UOKNOR

 Location/Qualifiers
 1. 203077
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /map="17"
 /clone="rp23-387d2"
 /clone_lib="RPCI - 23 Female (C57BL/6J) Mouse BAC Library"

 Query Match 53.8%; Score 666.6; DB 10; Length 203077;
 Best Local Similarity 89.8%; Pred. No. 1e-140;
 Matches 766; Conservative 0; Mismatches 69; Indels 18; Gaps 4;

 QY 387 GAACTGAATGACTTCATCCAGCTTGATGAAGAGTACAGGATGATGAGATGACCAT 446
 Db 135004 GAACTGAATGACTTCATCCATTCATCTGATGAAGAGTACAGGATGATGAGATGACCAT 134945
 QY 447 CTAGTTCTGACGAGCAGCTGAAGAGACAGCGCAGCTGCTCAGTACACGAGCAG 506
 Db 134944 CTAGTTCTTTCAGCAACAATTGAAGAGACAGCAGCAGCTGCTCAGTACACGAGCAG 134885
 QY 507 CAGTCTCAGGCTCTGCCCCAAGTACCAGCAGCTACAGCTTCTCAACTCTAGACAG 566
 Db 134884 CAGTCTCAGGCTCAGCTCCAGTACAGCAGCTACATCTTCTGAACCTGTAGATCAG 134825
 QY 567 TCAGAGGCCAAGTAAGACTCGAGTCTGCTGACAAACGACCAAGTAATGTAGCTCC 626
 Db 134824 GCAGAGGTCAACAGCAAGACTGCGTCTGCGCAATGGACCAAGTAATGCGAGCTCC 134765
 QY 627 TCCGCGCAGAGGCTGCTGGTCTGATTTACAGGGAGGGCAACACACCGAGATGAC 686
 Db 134764 TCCGCGCAGAGGAGCTGCTGGTCTGATTTACAGGGAGGGGAGCACCTCAGGATGAC 134705
 QY 687 TTTCCTTCTTCTCAGGGAAATGGTAATAAGTCTCCAAACAGCTCAGAGGAGAGAACTGGC 746
 Db 134704 TTTCCTTCTTCTCAGGGAAATGGTAATAAGGCTCCCAACAGCTCAGAGGAGAGAACTGGC 134645

QY 747 AGAGGAGGTAGTGGTTACGTAATCAACTCACTCAGTGGGGGTATGAAAGTGTAGACTCTCC 806
 Db 134644 AGAGGAGGTAGTGGTTACGTAATCAACTCACTCAGTGGGGGTATGAAAGTGTAGACTCTCC 134585
 QY 807 ACGGGAGGTGAAAGTCTCTCTCACACCAATCAAAATGACACAGCTCCAGTCTACACCT 866
 Db 134584 ACGGGAGGTGAAAGTCTCTCTCACACCAATCAAAATGACACAGCTCCAGTCTACACCT 134525
 QY 867 CAAGAGGAGAAAGCAGTGAAGTGGGAAAGTAATCGAACTGGGGTCCCGCCACGTTCCAG 926
 Db 134524 CAAGAGGAGAAAGCAGTGAAGTGGGAAAGTAATCGAACTGGGGTCCCGCCACGTTCCAG 134465
 QY 927 AATGGCTTGGACTCAAGTGTAAATGTACAGGGTTCAGTTCGTAATATTTTTCAGCAAA 986
 Db 134464 AATGGCTTGGACTCAAGTGTAAATGTACAGGGTTCAGTTCGTAATATTTTTCAGCAAA 134405
 QY 987 TTTTATACAGTGTCAATTAATTTGGGAGAGATCTCTCCAGAAAAATAATGCAATCTT 1046
 Db 134404 TTTTATACAGTGTCAATTAATTTGGGAGAGATCTCTCCAGAAAAATAATGCAATCTT 134346
 QY 1047 TTGTACCAATTTGCTTTTGTGGGTGACGTTTGTGGTTCGTTTGTGTTTGTCT 1106
 Db 134345 TTGTACCAATTTGCTTTTGTGGGTGACGTTTGTGGTTCGTTTGTGTTTGTCT 134288
 QY 1107 TTGGAATTTTCTTTTCTTTTAAAGAGGTGGTGTGTTTGTGTTTCTCAATCTTC 1166
 Db 134287 TTTTCTTTTCTTTTCTTTTAAAGAGGTGGTGTGTTTGTGTTTCTCAATCTTC 134242
 QY 1167 TCGCGCTTGGGAATGTAACAGTAAATTAATTTGAATGTTCC-AAAAGGCATTTGTG 1225
 Db 134241 TCGCACTTTGAAATGTAACAGTAAATTAATTTGAATGTTCC-AAAAGGCATTTGTG 134182
 QY 1226 GAGGCTCCAGTTA 1238
 Db 134181 TAGGGTCAAGTTA 134169

RESULT 10
 AL589878
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

AL589878 203949 bp DNA linear ROD 05-APR-2002
 Mouse DNA sequence from clone RP23-351C7 on chromosome 15, complete
 sequence.
 AL589878
 AL589878.12 GI:20068450
 HTG.
 Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1
 Phillimore,B.
 Direct Submission
 Submitted (04-APR-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On Apr 7, 2002 this sequence version replaced gi:16304732.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em: EMBL; Sw:
 SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP

database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-351C7 is
from the RPI-23 Mouse PAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6.

FEATURES

Location/Qualifiers
1..203949
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="15"
/clone="RP23-351C7"
/clone_lib="RPI-23"

ORIGIN

Query Match 53.8%; Score 666.6; DB 10; Length 203949;
Best Local Similarity 89.8%; Pred. No. 1e-140;
Matches 766; Conservative 0; Mismatches 69; Indels 18; Gaps 4;
387 GAACCTGAATGACTTCATCATCCAGCTTGATGAAGAAGTAGAGGGTATGCAGAGTACCATT 446
Db 68594 GAACCTGAATGACTTCATCATCCAGCTTGATGAAGAAGTAGAGGGTATGCAGAGTACCATT 68653
387 GAACCTGAATGACTTCATCATCCAGCTTGATGAAGAAGTAGAGGGTATGCAGAGTACCATT 446
Db 68594 GAACCTGAATGACTTCATCATCCAGCTTGATGAAGAAGTAGAGGGTATGCAGAGTACCATT 68653
447 CTAGTTCTGACGAGCAGCTGAAGGAGACACGCCAGCAGTTGGCTCAGTACCGAGCAGCAG 506
Db 68654 CTAGTTCTGACGAGCAGCTGAAGGAGACACGCCAGCAGTTGGCTCAGTACCGAGCAGCAG 506
507 CAGTCTCAGCGCTCTGCCCAAGTACCAGCAGGACTACAGCTTCTGAACCTGTAGAACACAG 566
Db 68714 CAGTCTCAGCGCTCTGCCCAAGTACCAGCAGGACTACAGCTTCTGAACCTGTAGAACACAG 566
567 TCAGAGGCCACAACTAAAGACTCGAGTCGTCTGACAAACGACCAAGTAATGTAGTCTCC 626
Db 68774 GCAGAGGTCAACAGCAAGACTCGAGTCGTCTGACAAACGACCAAGTAATGTAGTCTCC 68833
627 TCCCGCCAGAGGAGCTCTGGTCTGGATTTACAGAGGAGGAGCACAACCGAGAGTAC 686
Db 68834 TCCCGCCAGAGGAGCTCTGGTCTGGATTTACAGAGGAGGAGCACAACCGAGAGTAC 68893
687 TTTCCTTCTTCTCCAGGGAATGGTAATAAGTCTCTCCACAGCTCAGAGGAGAGAACTGGC 746
Db 68894 TTTCCTTCTTCTCCAGGGAATGGTAATAAGTCTCTCCACAGCTCAGAGGAGAGAACTGGC 68953
747 AGAGAGGTAGTGGTTAGTAAATCAACTCAGTCGGGGGTATGAAGTGTAGTCTCCC 806
Db 68954 AGAGAGGTAGTGGTTAGTAAATCAACTCAGTCGGGGGTATGAAGTGTAGTCTCCC 69013
807 ACGGCACTGAAACTCTCTCACACCAATCAATGACACAGACTCCAGTCTCATGACCT 866
Db 69014 ACGGCACTGAAACTCTCTCACACCAATCAATGACACAGACTCCAGTCTCATGACCT 69073
867 CAAGAGGAGAAAGCAGTGGGAAAGGTAAATCGAACTGGGGTTCGCCGACGTTTCAG 926
Db 69074 CAAGAGGAGAAAGCAGTGGGAAAGGTAAATCGAACTGGGGTTCGCCGACGTTTCAG 69133
927 AATGCTTGGACTCAGTGAATGATACAGGTTTCAGTTTGTATATTTTTCAGCAAA 986
Db 69134 AATGCTTGGACTCAGTGAATGATACAGGTTTCAGTTTGTATATTTTTCAGCAAA 69193
987 TTTTATACAGTGTCAATTAATTTGGGAGAGATACGTCCAGAAAATTAATGATACATT 1046
Db 69194 TTTTATACAGTGTCAATTAATTTGGGAGAGATACGTCCAGAAAATTAATGATACATT 69252
1047 TTGTCACAAATTTGCTTTTGGGTGTACGTTTGTGTTTTTTTGTGTTTTTTTCT 1106
Db 69253 TTGTCACAAATTTGCTTTTGGGTGTACGTTTGTGTTTTTTTGTGTTTTTTTCT 69310
1107 TTGGATTTTTTTTCTTTCTTTTAAAGAGGTGGTGTGTTTTTTTCTTCCTCAATCTTC 1166
Db 69311 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 108685
1167 TCGCGCTGGGAATGTGAAGTAACTTTGAAATGTTGC-AAAAGGCAATTTGTG 1225

69357 TGCCACTTTGAAAATTTCTAAACAGTTAATTACTTTTGAACATTTGCTAAAGGACATTTTGTG 69416
QY 1226 GAGGGTCCAGTTA 1238
Db 69417 TAGGGTCAAGTTA 69429
RESULT 11
AC074222.1/c
WPCOMMENT
Sequence split into 5 fragments LOCUS AC074222 Accession AC074222
Fragment Name Begin End
AC074222.0 1 110000
AC074222.1 100001 210000
AC074222.2 200001 310000
AC074222.3 300001 410000
AC074222.4 400001 412547
Continuation (2 of 5) of AC074222 from base 100001 (AC074222 Mus musculus chromosome 17
Query Match 52.9%; Score 654.6; DB 2; Length 110000;
Best Local Similarity 89.7%; Pred. No. 5.4e-138;
Matches 765; Conservative 0; Mismatches 69; Indels 19; Gaps 5;
387 GAACCTGAATGACTTCATCATCCAGCTTGATGAAGAAGTAGAGGGTATGCAGAGTACCATT 446
Db 109400 GAACCTGAATGACTTCATCATCCAGCTTGATGAAGAAGTAGAGGGTATGCAGAGTACCATT 109341
447 CTAGTTCTGACGAGCAGCTGAAGGAGACACGCCAGCAGTTGGCTCAGTACCGAGCAGCAG 506
Db 109340 CTAG-TCTCAGCAACAATTGAAGGAGACACGCCAGCAGTTGGCAGCAGTACCGAGCAG 109282
507 CAGTCTCAGCGCTCTGCCCAAGTACCAGCAGGACTACAGCTTCTGAACCTGTAGAACACAG 566
Db 109281 CAGTCTCAGCGCTCTGCCCAAGTACCAGCAGGACTACAGCTTCTGAACCTGTAGAACACAG 109222
567 TCAGAGGCCACAACTAAAGACTCGAGTCGTCTGACAAACGACCAAGTAATGTAGTCTCC 626
Db 109221 GCAGAGGTCAACAGCAAGACTCGAGTCGTCTGACAAACGACCAAGTAATGTAGTCTCC 109162
627 TCCCGCCAGAGGAGCTCTGGTCTGGATTTACAGAGGAGGAGCACAACCGAGAGTAC 686
Db 109161 TCCCGCCAGAGGAGCTCTGGTCTGGATTTACAGAGGAGGAGCACAACCGAGAGTAC 109102
687 TTTCCTTCTTCTCCAGGGAATGGTAATAAGTCTCTCCACAGCTCAGAGGAGAGAACTGGC 746
Db 109101 TTTCCTTCTTCTCCAGGGAATGGTAATAAGTCTCTCCACAGCTCAGAGGAGAGAACTGGC 109042
747 AGAGAGGTAGTGGTTAGTAAATCAACTCAGTCGGGGGTATGAAGTGTAGTCTCCC 806
Db 109041 AGAGAGGTAGTGGTTAGTAAATCAACTCAGTCGGGGGTATGAAGTGTAGTCTCCC 108982
807 ACGGCACTGAAACTCTCTCACACCAATCAATGACACAGACTCCAGTCTCATGACCT 866
Db 108981 ACGGCACTGAAACTCTCTCACACCAATCAATGACACAGACTCCAGTCTCATGACCT 108922
867 CAAGAGGAGAAAGCAGTGGGAAAGGTAAATCGAACTGGGGTTCGCCGACGTTTCAG 926
Db 108921 CAAGAGGAGAAAGCAGTGGGAAAGGTAAATCGAACTGGGGTTCGCCGACGTTTCAG 108862
927 AATGCTTGGACTCAGTGAATGATACAGGTTTCAGTTTGTATATTTTTCAGCAAA 986
Db 108861 AATGCTTGGACTCAGTGAATGATACAGGTTTCAGTTTGTATATTTTTCAGCAAA 108802
987 TTTTATACAGTGTCAATTAATTTGGGAGAGATACGTCCAGAAAATTAATGATACATT 1046
Db 108801 TTTTATACAGTGTCAATTAATTTGGGAGAGATACGTCCAGAAAATTAATGATACATT 108743
1047 TTGTCACAAATTTGCTTTTGGGTGTACGTTTGTGTTTTTTTGTGTTTTTTTCT 1106
Db 108742 TTGTCACAAATTTGCTTTTGGGTGTACGTTTGTGTTTTTTTGTGTTTTTTTCT 108685
1107 TTGGAATTTTTTCTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 1166


```

||||| 685 CTTGATGAGGAGGTAGAGGGTATGCGAGTACCAATCTAGTGCTTCAGCAGAGTGTGAAG 744
||||| 471 GAGACACGCGCAGAGTGGCTAGTACACGAGCAGCAGTCTCAGGCGCTCTGCCCAAGT 530
||||| 745 GAGACTCGCAGCAGTGGCGGAGTACACGAGCAGCAGTCCAGGCGCTCCACCCAGT 804
||||| 531 ACCAGCAGGACTACAGCTTCTGAACTGTAGAACAGTACAGAGCCACAGAGTAAAGACTGC 590
||||| 805 ACCAGCAGGAGTCCATCTTCTGAGCCTACAGACAGGAGAGGCTGTGGGTAAAGACTGC 864
||||| 591 AGTCGCTGACAAACGAGCAAGTAATGGTGTAGTCTCTCCGCGCAGAGGAGCTGTGGGTCT 650
||||| 865 AGCGCGCTGCTTACGAGCCACAGCAATGGTGTAGTCTCTCCATCAGCGGAGCTGTGGGTCT 924
||||| 651 GGATTTTACAGGAGGCGCAACACCAACGGAAGATGATCTTCTCTTCTCAGGGAATGCT 710
||||| 925 GGATTTTATAGGAGGAGGTAGCAGCAGCAAGAGATGATCTTCCGGCTTCTCCAGGGAATGCT 984
||||| 711 AATAAGTCTCCACAGCTCAGAGGAGAGCACTGGCAGAGGAGGTAGTGTACGTAAAT 770
||||| 985 AATAAGCTGTCCAAACCACTCTGAAGATAGAACTGGTAGAGAGGTGTAGTACATTAAC 1044
||||| 771 CAACTCAGTGGGGGTATGAAGTGTAGACTCTCCACGCGGCGAGTGAAGAACTCTTCACA 830
||||| 1045 CAACTCAGTACTGGGTATGAAGTGTAGACTCTCCCACTGGCAGTGAAGAACTCTCTCACT 1104
||||| 831 CACCAATCAATGACACAGACTCAGTCATGACCTCAGAGGAGAGCAAGCAGTGTGGG 890
||||| 1105 CACCATTCAATGACACAGACTCCCAATCATGATCATCTCCTCAAGAGGAGAGCAAGTGCATG 1164
||||| 891 AAGGTAATCGAAGTGGGTTCGCGCACGTTCCAGAAATGGCTTGGACTCAAGTGTAAAT 950
||||| 1165 AAGGTAACAGAACTGGGTTCTCGCATGTCCAGATGGTTGGACTCAATGTAAAT 1224
||||| 951 GTACAGGGTTCAGTTTGTATATTTTTCAGCAAAATTTTATACAGTGCATTTAATTT 1010
||||| 1225 GTGAGGGTTCAGTTTGTAACTTTTCTGCAAGTATTTTATACAGTGCATTTAATTT 1284
||||| 1011 GGGAGAGTACTCTCCAGA-----AAATTAATGATATCTTTTGTCACAATTTGC 1060
||||| 1285 GGGAGAGATACTCTTCAGAAAGCGTTATTAATAGTGCATTTGTCACAATTTTGC 1344
||||| 1061 CTTTTTGTGGTGTACGTTT 1080
||||| 1345 CTTTTTGTGGATTTCTTTT 1364

RESULT 14
AC112423 Rattus norvegicus clone CH230-29021, WORKING DRAFT SEQUENCE, 7
LOCUS AC112423 165096 bp DNA linear HTG 15-NOV-2002
DEFINITION Rattus norvegicus clone CH230-29021, WORKING DRAFT SEQUENCE, 7
unordered pieces.
ACCESSION AC112423 GI:25006897
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Rattus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 165096)
Muzny,D,Marie., Metzker,M, Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allien,H., Alébrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chu,J.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Cree,A., D'Souza,L.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,

```

Draper, H., Dugan-Rocha, S., Dunn, K., Durbin, K., Duval, B., Eaves, K., Bean, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, A., Carner, T., Garza, M., Gubert, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Loresuhewa, L., Loulseghe, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartine, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapus, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwackeleme, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Popper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, R., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Wang, V., Villanueva, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wood, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, R., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL
TITLE
JOURNAL
AUTHORS
COMMENT

Direct Submission
Unpublished
2 (bases 1 to 165096)

Worley, K.C.

Direct Submission

Submitted (21-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 165096)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 15, 2002 this sequence version replaced gi:23270045. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a contig-scaffold). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hqsc-help@bcm.tmc.edu

----- Project Information

Center project name: GRST

Center clone name: CH230-29021

47 AGAGGAGGAGTGGTTACGTAATCAACTCAGTGGGGGTATGAAAGTGTAGACTCTCCC 806


```
QY      867 CAAGAGGAGAAAGCAGTGTGGGAAAGGTAATCGAACTGTGGGTCCGCCACGTTCAAG 926
Db      193926 CAAGAGGAGAAAGCAGTGTGGGAAAGGTAATCGAACTGTGGGTCCGCCACGTTCAAG 193985
QY      927 AATGGCTTGGACTCAAGTGTAAATGTACAGGGTTCAGTTTTGTAAATTTTTTCAGCAAA 986
Db      193986 AATGGCTTGGACTCAAGTGTAAATGTACAGGGTTCAGTTTTGTAAATTTTTTCAGCAAA 194045
QY      987 TTTTATACAGTGTCAATTTAATTTGGGAGAGGATCTGTCCAGAAAATTAATGCATACTT 1046
Db      194046 TTTTATACAGTGTCAATTTAATTTGGGAGAGGATCTGTCCAGAAAATTAATGCATACTT 194103
QY      1047 TTGTCACAATTTGCCCTTTTGTGGGTGTACGTTTTGGTTTTTTTGTGTTTTTTTTCT 1106
Db      194104 TTGTCACAATTTATACCCCTTTTGTGGGTGTGTTTTGTGTTGGTCTTTTTTT--- 194160
QY      1107 TTGGAATTTTTTCTTTCTTTTCTTTTAAAGAGGTGGTGTGTTTTTGTCTTCAATCTTC 1166
Db      194161 -----TTTTTTTAAATGCTTCAATACCTC 194186
QY      1167 TCGCGCTTGGGAATGTAAACAGTTAATTAATCTTTGAATGTGC-AAAAGGCAATTTGTG 1225
Db      194187 TGCTACTTTGAAAATGTAAACAGTTAATTAATCTTTGAACATGTCTAAAAGGACATTTTGTG 194246
QY      1226 GAGGTCCAGTTA 1238
Db      194247 TAGGTCAGTTA 194259
```

Search completed: November 14, 2004, 12:54:06
Job time : 5531 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2004, 11:21:58 ; Search time 1973 Seconds
(without alignments)
3388.962 Million cell updates/sec

Title: US-10-030-389-12

Perfect score: 1238

Sequence: 1 acatcttggaatgcgacta.....ttttgtgagggtccagta 1238

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3625171 seqs, 2700493622 residues

Total number of hits satisfying chosen parameters: 7250342

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:
18: /cgn2_6/ptodata/1/pubpna/US10F_NEW_PUB.seq:
19: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:
20: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:
21: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1116.6	90.2	2772	15	US-10-252-157-398
2	797.4	64.4	41309	17	US-10-741-601-5771
3	379.8	30.7	31657	17	US-10-741-601-5690
4	329.8	26.6	12648	17	US-10-741-601-5652
5	296.6	24.0	440	9	US-09-983-965-217
6	226.2	18.3	480	16	US-10-242-535A-15856
7	226.2	18.3	480	16	US-10-085-783A-15856
8	226.2	18.3	1600	9	US-09-840-787-50
9	226.2	18.3	1600	13	US-10-044-090-550
10	226.2	18.3	1622	9	US-09-954-456-492
11	226.2	18.3	1622	15	US-10-172-118-1100
12	226.2	18.3	1622	16	US-10-342-887-1100

13	226.2	18.3	1691	13	US-10-044-090-552	Sequence 552, App
14	226.2	18.3	1996	13	US-10-044-090-551	Sequence 551, App
15	207.6	16.8	694	18	US-10-425-115-6885	Sequence 6885, Ap
16	201.2	16.3	475	9	US-09-867-701-2417	Sequence 2417, Ap
17	200.6	16.2	201	17	US-10-741-601-25687	Sequence 25687, A
18	191.4	15.5	601	14	US-10-066-543-1379	Sequence 1379, Ap
19	150.6	12.2	201	17	US-10-741-601-25688	Sequence 25688, A
20	143.4	11.6	3902	13	US-10-044-090-451	Sequence 451, App
21	143.4	11.6	3902	15	US-10-084-817-260	Sequence 260, App
22	134.8	10.9	201	17	US-10-741-601-14585	Sequence 14585, A
23	134.8	10.9	201	17	US-10-741-601-25705	Sequence 25705, A
24	120.8	9.8	2642	16	US-10-108-260A-2162	Sequence 2162, Ap
25	120.8	9.8	81826	14	US-10-175-523-197	Sequence 197, App
26	115.2	9.3	501	9	US-09-833-790-179	Sequence 179, App
27	108	8.7	481	10	US-09-918-395-29597	Sequence 29597, A
28	93	7.5	558	9	US-09-728-446-797	Sequence 797, App
29	75	6.1	501	9	US-09-833-790-168	Sequence 168, App
30	75	6.1	33053	17	US-10-433-793-35	Sequence 35, Appl
31	72.4	5.8	6794	16	US-10-321-613-133	Sequence 133, App
32	71.4	5.8	40324	17	US-10-433-793-179	Sequence 179, App
33	69	5.6	5823	15	US-10-340-453-256	Sequence 256, App
34	68.5	5.5	785	17	US-10-767-701-8312	Sequence 8312, App
35	63	5.5	12007	15	US-10-311-455-690	Sequence 690, App
36	67	5.4	201	17	US-10-741-601-25715	Sequence 25715, A
37	67	5.4	18624	15	US-10-311-455-1676	Sequence 1676, Ap
38	66.3	5.4	5413	16	US-10-221-714A-418	Sequence 418, App
39	66.5	5.4	411	9	US-09-960-352-14521	Sequence 14521, A
40	66.5	5.4	447	18	US-10-425-115-148961	Sequence 148961,
41	66.5	5.4	6668	15	US-10-311-455-1669	Sequence 1669, Ap
42	66.5	5.4	3673778	15	US-10-312-841-1	Sequence 1, Appli
43	66.4	5.4	516	17	US-10-021-323-16996	Sequence 16996, A
44	66.2	5.3	554	18	US-10-425-115-139665	Sequence 139665,
45	66.2	5.3	1287	16	US-10-424-599-97419	Sequence 97419, A

ALIGNMENTS

RESULT 1

US-10-252-157-398
Sequence 398, Application US/10252157
Publication No. US20030190640A1
GENERAL INFORMATION:
APPLICANT: Paris, Mary
TITLE OF INVENTION: GENES EXPRESSED IN PROSTATE CANCER
FILE REFERENCE: PA-0027-1 US
CURRENT APPLICATION NUMBER: US/10/252,157
CURRENT FILING DATE: 2002-10-01
PRIOR APPLICATION NUMBER: 60/295,048
PRIOR FILING DATE: 2001-05-31
NUMBER OF SEQ ID NOS: 501
SOFTWARE: PERL Program
SEQ ID NO 398
LENGTH: 2772
TYPE: DNA
ORGANISM: Homo sapiens
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20030190640A1 040057.2
US-10-252-157-398

Query Match 90.2%; Score 1116.6; DB 15; Length 2772;
Best Local Similarity 99.2%; Pred. No. 3.2e-265;
Matches 1122; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 1 ACATCTTGGATGCGACTAGCAACCAAGGCAAGATATTCAAGAGTGTTACTACTCAA 60
1641 ACATCTTGGATGCGACTAGCAACCAAGGCAAGATGCGAGAGTGTTACTACTCAA 1700
Db
QY 61 TCCAGTACTCAAGCAAGTCAGCAGCCTAGCGTGTGCCCACTTAGTATCAACAATGCTAG 120
1701 TCCAGTACTCAAGCAAGTCAGCAGCCTAGCGTGTGCCCACTGAGATCAACAATGCTAG 1760

```

QY 121 ACCAGCGATCAACTGTGTTTTCCTAAATAAGTGAAGTGAACAGACTAAAGACA 180
DB 1761 ACCAGCGATCAACTGTGTTTTCCTAAATAAGTGAAGTGAACAGACTAAAGACA 1820
QY 181 AACTGGAACAGCCCAAAATGAAGTGGCTGGAAGTTTACGCCCTGATAGCCAAACAG 240
DB 1821 AACTGGAACAGCCCAAAATGAAGTGGCTGGAAGTTTACGCCCTGATAGCCAAACAG 1880
QY 241 GGAATAAGTTAATGCGAAGTGTGCAATGCTTATCCAGGAGAAATCAAGAGCTTGAAGGC 300
DB 1881 GGAATAAGTTAATGCGAAGTGTGCAATGCTTATCCAGGAGAAATCAAGAGCTTGAAGGC 1940
QY 301 AGCTGTCCAGGAGCTATTCACAACTTGAGCAGAGTTGGCTTTACAGAAAGAAATACA 360
DB 1941 AGCTGTCCAGGAGCTATTCACAACTTGAGCAGAGTTGGCTTTACAGAAAGAAATACA 2000
QY 361 GTGAGGAGCTTAAAGCAGTCAAGTGAATGAATGACTTTCATCATCCAGCTTGCATGAAG 420
DB 2001 GTGAGGAGCTTAAAGCAGTCAAGTGAATGAATGACTTTCATCATCCAGCTTGCATGAAG 2060
QY 421 AAGTAGAGGCTATGCAGAGTACCAATTCAGTTCTGACAGCAGCTGAAGAGACACGCC 480
DB 2061 AAGTAGAGGCTATGCAGAGTACCAATTCAGTTCTGACAGCAGCTGAAGAGACACGCC 2120
QY 481 AGCAGTTGGCTCAGTACCAGCAGCAGCTCTCAGGCTCTGCCCAAGTACCAGCAGGA 540
DB 2121 AGCAGTTGGCTCAGTACCAGCAGCAGCTCTCAGGCTCTGCCCAAGTACCAGCAGGA 2180
QY 541 CTACAGCTTCTGAACCTGTGAAACAGTCAAGGAGCCCAAGTAAAGACTGCACTCGTCTGA 600
DB 2181 CTACAGCTTCTGAACCTGTGAAACAGTCAAGGAGCCCAAGTAAAGACTGCACTCGTCTGA 2240
QY 601 CAACCGGACCAAGTAATGGTAGCTCTCCCGCAGAGGAGCTCTGGGTCTGGATTTTACA 660
DB 2241 CAACCGGACCAAGTAATGGTAGCTCTCCCGCAGAGGAGCTCTGGGTCTGGATTTTACA 2300
QY 661 GGGAGGGCAACACAAACGAGATGACTTTCCTTCTCTCCAGGGAATGGTAATAGTCCT 720
DB 2301 GGGAGGGCAACACAAACGAGATGACTTTCCTTCTCTCCAGGGAATGGTAATAGTCCT 2360
QY 721 CCACAGCTCAGAGGAGAGTGGCAGAGAGGTAGTGTAGCTTAAGTCAACTCAGTG 780
DB 2361 CCACAGCTCAGAGGAGAGTGGCAGAGAGGTAGTGTAGCTTAAGTCAACTCAGTG 2420
QY 781 CGGGGTATGAAGTGTAGACTCTCCACCGGCGAGTGAATACTCTCTCACACCAATCAA 840
DB 2421 CGGGGTATGAAGTGTAGACTCTCCACCGGCGAGTGAATACTCTCTCACACCAATCAA 2480
QY 841 ATGACAGAGTCCAGTCACTGACCTCAAGAGGAGAAAGCAGTGAAGTGAAGTAATC 900
DB 2481 ATGACAGAGTCCAGTCACTGACCTCAAGAGGAGAAAGCAGTGAAGTGAAGTAATC 2540
QY 901 GAACTGTGGGTTCGCGCCACGTTTCAAGATGGCTTGGACTCAAGTGAATAATGACAGGTT 960
DB 2541 GAACTGTGGGTTCGCGCCACGTTTCAAGATGGCTTGGACTCAAGTGAATAATGACAGGTT 2600
QY 961 CAGTTTGTAAATATTTTTCAGCAAAATTTTATACAGTGTCAATTTAAATTTGGAGAGAT 1020
DB 2601 CAGTTTGTAAATATTTTTCAGCAAAATTTTATACAGTGTCAATTTAAATTTGGAGAGAT 2660
QY 1021 ACTGTCCAGAAATTAATGCATCTTTTGTCAAAATTTTCCCTTTTGTGGGTGTAGCTT 1080
DB 2661 ACTGTCCAGAAATTAATGCATCTTTTGTCAAAATTTTCCCTTTTGTGGGTGTAGCTT 2720
QY 1081 TGGTTTTTTTGTGTTTTTTTCTTTGGATTTTTTTTCTTTCTTTTTT 1131
DB 2721 TGGTTTTTTTGTGTTTTTTTCTTTGGATTTTTTTTCTTTCTTTTTT 2771

```

RESULT 2

US-10-741-601-5771

: Sequence 5771, Application US/10741601

```

; Publication No. US20040166519A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE OF INVENTION: STENOSIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001500
; CURRENT APPLICATION NUMBER: US/10/741,601
; NUMBER OF SEQ ID NOS: 26415
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5771
; LENGTH: 41309
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-741-601-5771

```

```

Query Match 64.4%; Score 797.4; DB 17; Length 41309;
Best Local Similarity 97.4%; Pred. No. 1e-185;
Matches 831; Conservative 1; Mismatches 17; Indels 4; Gaps 2;

QY 387 GAACCTGAATGACTTTCATCATCCAGCTTGTGAAGAGTAGAGGTTATGCAGAGTACCACTT 446
DB 34018 GAACCTGAATGACTTTCATCATCCAGCTTGTGAAGAGTAGAGGTTATGCAGAGTACCACTT 34077
QY 447 CTAGTTCTGCAGCAGCAGCTGAGGAGACACGCCAGCAGTTGGCTCAGTACAGCAGCAG 506
DB 34078 CTAGTTCTGCAGCAGCAGCTGAGGAGACACGCCAGCAGTTGGCTCAGTACAGCAGCAG 34137
QY 507 CAGTCTCAGGCTCTGCCCTCCCAAGTACCAGCAGGACTCAGCTTCTGAACCTGTAGAACAG 566
DB 34138 CAGTCTCAGGCTCTGCCCTCCCAAGTACCAGCAGGACTCAGCTTCTGAACCTGTAGAACAG 34197
QY 567 TCAGAGGCCACAAAGTAAAGACTGCGTCTGTAACCGGACCAAGTAATGTTAGTCTCC 626
DB 34198 TCAGAGGCCACAAAGTAAAGACTGCGTCTGTAACCGGACCAAGTAATGTTAGTCTCC 34257
QY 627 TCCGCCCAGAGAGCTCTGGGCTCTGGATTTCACAGGGAGGGGCAACAAACCGAGAGTAC 686
DB 34258 TCCGCCCAGAGAGCTCTGGGCTCTGGATTTCACAGGGAGGGGCAACAAACCGAGAGTAC 34317
QY 687 TTTCTCTCTTCTCCAGGGAATGGTAATAGTCTTCAACAGCTCAGAGGAGAGAACTGGC 746
DB 34318 TTTCTCTCTTCTCCAGGGAATGGTAATAGTCTTCAACAGCTCAGAGGAGAGAACTGGC 34377
QY 747 AGAGAGGTAGTGTGTTAAGTAAATCAACTCAGTCCGGGTTATGAAGTGTAGACTCTCCC 806
DB 34378 AGAGAGGTAGTGTGTTAAGTAAATCAACTCAGTCCGGGTTATGAAGTGTAGACTCTCCC 34437
QY 807 ACGGCGAGTGAATACTCTCTCACACCAATCAATGACACAGACTCCAGTCCATGACCCCT 866
DB 34438 ACGGCGAGTGAATACTCTCTCACACCAATCAATGACACAGACTCCAGTCCATGACCCCT 34497
QY 867 CAAGAGGAGAAAGCAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 926
DB 34498 CAAGAGGAGAAAGCAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 34557
QY 927 AATGGCTTGGACTCAAGTGAATAATGTACAGGTTTCAAGTGTAGTTTGTAAATTTTTCAGAAA 986
DB 34558 AATGGCTTGGACTCAAGTGAATAATGTACAGGTTTCAAGTGTAGTTTGTAAATTTTTCAGAAA 34617
QY 987 TTTTATACAGTGTCAATTAATTTGGGAGAGGATACGTCTCCAGAAATTAATGCACTT 1046
DB 34618 TTTTATACAGTGTCAATTAATTTGGGAGAGGATACGTCTCCAGAAATTAATGCACTT 34677
QY 1047 TTGTCAAAATTTGGCTTTTGTGGGTGAGCTTTTGTGTTTTTGTGTTTTTGTGTTTTTGT 1106
DB 34678 TTGTCAAAATTTGGCTTTTGTGGGTGAGCTTTTGTGTTTTTGTGTTTTTGTGTTTTTGT 34737
QY 1107 TTGATTTTTTTTCTTTTCTTTTAAAGAGGTGGTTTGTGTTTTTGTGTTTTTGTGTTTTTGT 1166
DB 34738 TTGATTTTTTTTCTTTTCTTTTAAAGAGGTGGTTTGTGTTTTTGTGTTTTTGTGTTTTTGT 34794
QY 1167 TGCAGCTTGGGAATGTGAACAGTTAATTTGATGTTGTC-AAAAGGGCATTTTGTG 1225

```

Db 34795 TCCCGCTTTGGAAATGTAAACAGTTAAATTTACATTGCTTAAAGGACATTTTGTG 34854
QY 1226 GAGGCTCCAGTTA 1238
Db 34855 TAGGGTCAAGTTA 34867

RESULT 3

US-10-741-601-5690
; Sequence 5690, Application US/10741601
; Publication No. US20040166519A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; STENOSIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001500
; CURRENT APPLICATION NUMBER: US/10/741,601
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 26415
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 5690
; LENGTH: 31657
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-741-601-5690

Query Match 30.7%; Score 379.8; DB 17; Length 31657;
Best Local Similarity 95.3%; Pred. No. 1.4e-82;
Matches 414; Conservative 0; Mismatches 17; Indels 4; Gaps 2;

QY 805 CCACGGGCGAGTGAAGAACTCTCTACACACCAATCAATGACACAGACTCCAGTCATGACC 864
Db 1 CCACGGGCGAGTGAAGAACTCTCTACACACCAATCAATGACACAGACTCCAGTCATGACC 60
QY 865 CTCAGAGGAGAAAGCAGTGCAGTGGGAAAGGTAAATCGAACTGTGGGTCCCGCCACGTTT 924
Db 61 CTCAGAGGAGAAAGCAGTGCAGTGGGAAAGGTAAATCGAACTGTGGGTCCCGCCACGTTT 120
QY 925 AGAATGCTTGGACTCAAGTGTAAATGTAACAGGTTTCAAGTTTGTAAATTTTTCAGCA 984
Db 121 AGAATGCTTGGACTCAAGTGTAAATGTAACAGGTTTCAAGTTTGTAAATTTTTCAGCA 180
QY 985 AATTTTATACAGTGCATTTAATTTGGGAGAGGATCTCCAGAGAAATTAATGATAC 1044
Db 181 AATTTTATACAGTGCATTTAATTTGGGAGAGGATCTCCAGAGAAATTAATGATAC 240
QY 1045 TTTTGTCAAAATTTGCCCTTTTGTGGGTGACGTTTGGTTTGTGTTTGTGTTTGT 1104
Db 241 TTTTGTCAAAATTTGCCCTTTTGTGGGTGACGTTTGGTTTGTGTTTGTGTTTGT 300
QY 1105 CTTTGGATTTTTCCTTTTCTTTTAAAGAGGTTGTTGTTTGTGTTTGTGTTTGT 1164
Db 301 CTTTGTGTTTTCCTTTTCTTTT---TTTTTTTTCCTTTTGTGTTTGTGTTTGT 357
QY 1165 TCTGGCGCTTGGAAATGTAAACAGTTAAATTTACATTTGAATTTGC-AAAAGGCAATTTG 1223
Db 358 TCTGGCGCTTGGAAATGTAAACAGTTAAATTTACATTTGAATTTGCTTAAAGGCAATTTG 417
QY 1224 TGGAGGGTCCAGTTA 1238
Db 418 TGTAGGGTCAAGTTA 432

RESULT 4

US-10-741-601-5652/c
; Sequence 5652, Application US/10741601
; Publication No. US20040166519A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; STENOSIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001500

; CURRENT APPLICATION NUMBER: US/10/741,601
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 26415
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 5652
; LENGTH: 12648
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-741-601-5652

Query Match 26.6%; Score 329.8; DB 17; Length 12648;
Best Local Similarity 94.5%; Pred. No. 2e-70;
Matches 364; Conservative 0; Mismatches 17; Indels 4; Gaps 2;

QY 855 AGTCATGACCTCAAGAGGAAAGCAGTGCAGTGGGAAAGGTAATCGAACTGTGGGTTC 914
Db 12648 AGTCATGACCTCAAGAGGAAAGCAGTGCAGTGGGAAAGGTAATCGAACTGTGGGTTC 12589
QY 915 CGCCACGTTTCAGAAATGGCTGGACTCAAGTGTAAATGTACAGGTTTCAGTTTGTAAAT 974
Db 12588 CGCCACGTTTCAGAAATGGCTGGACTCAAGTGTAAATGTACAGGTTTCAGTTTGTAAAT 12529
QY 975 TTTTTCAGCAAAATTTTATACAGTGTCAATTTAAATTTGGAGAGGATCTGTCAGAAAT 1034
Db 12528 TTTTTCAGCAAAATTTTATACAGTGTCAATTTAAATTTGGAGAGGATCTGTCAGAAAT 12469
QY 1035 TAATGCCATCTTTGTGCACAAATTTGCCCTTTTGTGGGTGTACGTTTGGTTTTTGT 1094
Db 12468 TAATGCCATCTTTGTGCACAAATTTGCCCTTTTGTGGGTGTACGTTTGGTTTTTGT 12409
QY 1095 TGTGTTTTTCTTTGGATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1154
Db 12408 TGTGTTTTTCTTTGGATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 12352
QY 1155 CTTCAATCTCTCGCGCTTGGGAAATTTGTAACAGTTAAATTTGAATTTGCT-AAAA 1213
Db 12351 CTTCAATCTCTCGCGCTTGGGAAATTTGTAACAGTTAAATTTGAATTTGCTTAAAT 12292
QY 1214 GGGCAATTTTGTGGAGGTTCCAGTTA 1238
Db 12291 GGCACATTTTGTAGGGTCAAGTTA 12267

RESULT 5

US-09-983-965-217
; Sequence 217, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Rao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 217
; LENGTH: 440
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 26-BOVMS1-017-Q1-E1-G5
US-09-983-965-217

Query Match 24.0%; Score 296.6; DB 9; Length 440;
Best Local Similarity 89.5%; Pred. No. 5.9e-63;
Matches 332; Conservative 0; Mismatches 34; Indels 5; Gaps 1;

QY 742 CTGGCAGAGGAGGTAGTGGTTACGTAAATCAACTCAGTGGCGGCTATGAAGTGTAGACT 801
 Db 1 CTGGCAGAGGAGGTAGTGGTTACGTAAATCAACTCAGTGGCGGCTATGAAGTGTAGACT 60
 QY 802 CTCACCGGCGAGTGAAGAACTCTCTCACACACCAATCAATGACACAGACTCCAGTCAATG 861
 Db 61 CTCACCGGCGAGTGAAGAACTCTCTCACACACCAATCAATGACACAGACTCCAGTCAATG 120
 QY 862 ACCCTCAAGAGCAGAGCAGTGGTGAAGGTAAATCGAATGTCGGTCCCGCCACG 921
 Db 121 ACCCGCAGAGAGAACTCTGAGTGGGAAGGTAAACCGAAGTCCGCGCTCCCGCCACG 180
 QY 922 TTCAGATGGTGGACTCAAGTGTAAATGACAGGGTTCAGTTTTGT-----AATATTT 976
 Db 181 TTCAGATGGTGGACTCAAGTGTAAATGACAGGGTTCAGTTTTGTAAAAAATAAT 240
 QY 977 TTCAGCAAAATTTATACAGTGTCAATTAATTTGGGAGGAGTACTGTCCAGAAATTA 1036
 Db 241 TTCAGCAAAATTTATACAGTGTCAATTAATTTGGGAGGAGTACTGTCCAGTAAATTA 300
 QY 1037 ATGCATATTTTGTCAAAATTTGCTTTTGTGGGTGTAGCTTTTGTGTTTTTTTGTG 1096
 Db 301 ATGCATATTTTGTCAAAATTTGCTTTTGTGGGTGTAGCTTTTGTGTTTTTTTGTG 360
 QY 1097 TTTTTCCTTT 1107
 Db 361 TTTCTTTGCTT 371

RESULT 6
 US-10-242-535A-15856
 ; Sequence 15856, Application US/10242535A
 ; Publication No. US20040013663A1
 ; GENERAL INFORMATION:
 ; APPLICANT: ChondroGene Inc.
 ; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
 ; FILE REFERENCE: 4231/2005
 ; CURRENT APPLICATION NUMBER: US/10/242,535A
 ; CURRENT FILING DATE: 2002-09-12
 ; PRIOR APPLICATION NUMBER: US 10/085,783
 ; PRIOR FILING DATE: 2002-02-28
 ; PRIOR APPLICATION NUMBER: US 60/305,340
 ; PRIOR FILING DATE: 2001-07-13
 ; PRIOR APPLICATION NUMBER: US 60/275,017
 ; PRIOR FILING DATE: 2001-03-12
 ; PRIOR APPLICATION NUMBER: US 60/271,955
 ; PRIOR FILING DATE: 2001-02-28
 ; NUMBER OF SEQ ID NOS: 58994
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 15856
 ; LENGTH: 480
 ; TYPE: DNA
 ; ORGANISM: Human
 US-10-242-535A-15856

Query Match 18.3%; Score 226.2; DB 16; Length 480;
 Best Local Similarity 96.7%; Pred. No. 1.5e-45;
 Matches 231; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
 QY 1 ACATCTTGGAAATCGACTAGCAACCAAGGACAAAGATATTCAGAGTGTACTACTCAA 60
 Db 51 ACATCTTGTAAATCGGACTAGCAACCAAGGACAAAGATATTCAGAGTGTACTACTCAA 110
 QY 61 TCAGTACTCTCAAGCAAGTCCAGCAGCGGTGCCCCAATAGATCAACAAATGGTAG 120
 Db 111 TCAGTACTCTCAAGCAAGTCCAGCAGCGGTGCCCCAATAGATCAACAAATGGTAG 170
 QY 121 ACCAGCGATCAACTGTTTCTTCTTAAATGAAGGTGAAGTGGAAACAGACTAAAGACA 180
 Db 171 ACCAGCGATCAACTGTTTCTTCTTAAATGAAGGTGAAGTGGAAACAGACTAAAGACA 230

QY 181 AACTGGAACAAGCCCAAAATGAATGAGTGCCTGGAAGTTTACGCTGATAGCCAAACA 239
 Db 231 AACTGGAACAAGCCCAAAATGAATGAGTGCCTGGAAGTTTACGCTGATAGCCAAACA 289
 RESULT 7
 US-10-085-783A-15856
 ; Sequence 15856, Application US/10085783A
 ; Publication No. US20040037841A1
 ; GENERAL INFORMATION:
 ; APPLICANT: ChondroGene Inc.
 ; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
 ; FILE REFERENCE: 4231/2002
 ; CURRENT APPLICATION NUMBER: US/10/085,783A
 ; CURRENT FILING DATE: 2002-02-28
 ; PRIOR APPLICATION NUMBER: US 60/305,340
 ; PRIOR FILING DATE: 2001-07-13
 ; PRIOR APPLICATION NUMBER: US 60/275,017
 ; PRIOR FILING DATE: 2001-03-12
 ; PRIOR APPLICATION NUMBER: US 60/271,955
 ; PRIOR FILING DATE: 2001-02-28
 ; NUMBER OF SEQ ID NOS: 58994
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 15856
 ; LENGTH: 480
 ; TYPE: DNA
 ; ORGANISM: Human
 US-10-085-783A-15856

Query Match 18.3%; Score 226.2; DB 16; Length 480;
 Best Local Similarity 96.7%; Pred. No. 1.5e-45;
 Matches 231; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
 QY 1 ACATCTTGGAAATCGACTAGCAACCAAGGACAAAGATATTCAGAGTGTACTACTCAA 60
 Db 51 ACATCTTGTAAATCGGACTAGCAACCAAGGACAAAGATATTCAGAGTGTACTACTCAA 110
 QY 61 TCAGTACTCTCAAGCAAGTCCAGCAGCGGTGCCCCAATAGATCAACAAATGGTAG 120
 Db 111 TCAGTACTCTCAAGCAAGTCCAGCAGCGGTGCCCCAATAGATCAACAAATGGTAG 170
 QY 121 ACCAGCGATCAACTGTTTCTTCTTAAATGAAGGTGAAGTGGAAACAGACTAAAGACA 180
 Db 171 ACCAGCGATCAACTGTTTCTTCTTAAATGAAGGTGAAGTGGAAACAGACTAAAGACA 230
 QY 181 AACTGGAACAAGCCCAAAATGAATGAGTGCCTGGAAGTTTACGCTGATAGCCAAACA 239
 Db 231 AACTGGAACAAGCCCAAAATGAATGAGTGCCTGGAAGTTTACGCTGATAGCCAAACA 289

RESULT 8
 US-09-840-787-50
 ; Sequence 50, Application US/09840787
 ; Patent No. US20020058264A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lal, Preeti
 ; Hillman, Jennifer L.
 ; Bandman, Olga
 ; Shah, Purvi
 ; Au-Young, Janice
 ; Yue, Henry
 ; Guegler, Karl J.
 ; Corley, Neil C.
 ; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
 ; NUMBER OF SEQUENCES: 98
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.
 ; STREET: 3174 Porter Drive
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94304

Query Match	18.3%;	Score 226.2;	DB 13;	Length 1600;	
Best Local Similarity	96.7%;	Pred. No. 2.9e-45;			
Matches 231;	Conservative	0;	Mismatches 8;	Indels 0;	Gaps 0;
QY	1	ACATCCTTGGAAATGCGACTAGCAACCAAGGGAACAAGATATTCAAGAGTGTACTACTCAA	60		
Db	311	ACATCCTTGTAAATGCGACTAGCAACCAAGGGAACAAGATGCAAGAGTGTACTACTCAA	370		
QY	61	TCCAGTACCTCAAGCAAGTCCAGACGCTTAGGTTGCCCACTTAGATCAACAATGGTAG	120		
Db	371	TCCAGTACCTCAAGCAAGTCCAGACGCGAGGTTGCCCACTAGATCAACAATGGTAG	430		
QY	121	ACCCAGAGGATCAACTTGTTTTCCTAAAAATGAAAGGTTGAATGCGAAACAGACTAAAGACA	180		
Db	431	ACCCAGAGGATCAACTTGTTTTCCTAAAAATGAAAGGTTGAATGCGAAACAGACTAAAGACA	490		
QY	181	AACTGGAAACCAAGCCCAAAATGAATGAATGAGTGCCTCGGAAGTTTACGGCTGATGCCAAACA	239		
Db	491	AACTGGAAACCAAGCCCAAAATGAATGAATGAGTGCCTCGGAAGTTTACGGCTGATGCCAAACA	549		
RESULT 10					
US-09-954-456-492					
; Sequence 492, Application US/09954456					
; Patent No. US20020115057A1					
; GENERAL INFORMATION:					
; APPLICANT: Young, Paul					
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents					
; FILE REFERENCE: 689290-76					
; CURRENT APPLICATION NUMBER: US/09/954,456					
; CURRENT FILING DATE: 2001-09-18					
; PRIOR APPLICATION NUMBER: US/60/233,617					
; PRIOR FILING DATE: 2000-09-18					
; PRIOR APPLICATION NUMBER: US/60/234,052					
; PRIOR FILING DATE: 2000-09-20					
; PRIOR APPLICATION NUMBER: US/60/234,923					
; PRIOR FILING DATE: 2000-09-25					
; PRIOR APPLICATION NUMBER: US/60/235,134					
; PRIOR FILING DATE: 2000-09-25					
; PRIOR APPLICATION NUMBER: US/60/235,637					
; PRIOR FILING DATE: 2000-09-26					
; PRIOR APPLICATION NUMBER: US/60/235,638					
; PRIOR FILING DATE: 2000-09-26					
; PRIOR APPLICATION NUMBER: US/60/235,711					
; PRIOR FILING DATE: 2000-09-27					
; PRIOR APPLICATION NUMBER: US/60/235,720					
; PRIOR FILING DATE: 2000-09-27					
; PRIOR APPLICATION NUMBER: US/60/235,840					
; PRIOR FILING DATE: 2000-09-27					
; PRIOR APPLICATION NUMBER: US/60/235,863					
; PRIOR FILING DATE: 2000-09-27					
; NUMBER OF SEQ ID NOS: 2276					
; SOFTWARE: PatentIn version 3.0					
; SEQ ID NO 492					
; LENGTH: 1622					
; TYPE: DNA					
; ORGANISM: Homo sapiens					
US-09-954-456-492					
Query Match	18.3%;	Score 226.2;	DB 9;	Length 1622;	
Best Local Similarity	96.7%;	Pred. No. 2.9e-45;			
Matches 231;	Conservative	0;	Mismatches 8;	Indels 0;	Gaps 0;
QY	1	ACATCCTTGGAAATGCGACTAGCAACCAAGGGAACAAGATATTCAAGAGTGTACTACTCAA	60		
Db	345	ACATCCTTGTAAATGCGACTAGCAACCAAGGGAACAAGATGCAAGAGTGTACTACTCAA	404		
QY	61	TCCAGTACCTCAAGCAAGTCCAGACGCTAGGTTGCCCACTTAGATCAACAATGGTAG	120		
Db	405	TCCAGTACCTCAAGCAAGTCCAGACGCGAGGTTGCCCACTAGATCAACAATGGTAG	464		

QY 121 ACCGAGCATCAACTGTTTTCCTAAATAATGAAAGGTGAAGTGAACAGACTAAAGACA 180
| | | | |
Db 465 ACCGAGCATCAACTGTTTTCCTAAATAATGAAAGGTGAAGTGAACAGACTAAAGACA 524
| | | | |
QY 181 AACTGGAAACAGCCCAAAATGAATGAGTGGCTGGAAGTTTACGCCTGATAGCCAAACA 239
| | | | |
Db 525 AACTGGAAACAGCCCAAAATGAATGAGTGGCTGGAAGTTTACGCCTGATAGGTAAACA 583
| | | | |
RESULT 11
US-10-172-118-1100
; Sequence 1100, Application US/10172118
; Publication No. US20030224374A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van de Vijver, Marc
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-999
; CURRENT APPLICATION NUMBER: US/10/172,118
; CURRENT FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/380,770
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 1100
; LENGTH: 1622
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM_004906
; DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-1100

Query Match 18.3%; Score 226.2; DB 15; Length 1622;
Best Local Similarity 96.7%; Pred. No. 2.9e-45;
Matches 231; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 1 ACATCCTTGGATCGGACTAGCAACCAAGGACAGATATTCAGAGTGTACTCTAAA 60
| | | | |
Db 345 ACATCCTTGGATCGGACTAGCAACCAAGGACAGATATTCAGAGTGTACTCTAAA 404
| | | | |
QY 61 TCCAGTACCTCAAGCAAGTCCAGCAGCTAGCGTTGCCCAACTTAGATCAACAATGGTAG 120
| | | | |
Db 405 TCCAGTACCTCAAGCAAGTCCAGCAGCTAGCGTTGCCCAACTTAGATCAACAATGGTAG 464
| | | | |
QY 121 ACCGAGCATCAACTGTTTTCCTAAATAATGAAAGGTGAAGTGAACAGACTAAAGACA 180
| | | | |
Db 465 ACCGAGCATCAACTGTTTTCCTAAATAATGAAAGGTGAAGTGAACAGACTAAAGACA 524
| | | | |
QY 181 AACTGGAAACAGCCCAAAATGAATGAGTGGCTGGAAGTTTACGCCTGATAGCCAAACA 239
| | | | |
Db 525 AACTGGAAACAGCCCAAAATGAATGAGTGGCTGGAAGTTTACGCCTGATAGGTAAACA 583
| | | | |

RESULT 12
US-10-342-887-1100
; Sequence 1100, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients

; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 1100
; LENGTH: 1622
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-1100
Query Match 18.3%; Score 226.2; DB 16; Length 1622;
Best Local Similarity 96.7%; Pred. No. 2.9e-45;
Matches 231; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 1 ACATCCTTGGATCGGACTAGCAACCAAGGACAGATATTCAGAGTGTACTCTAAA 60
| | | | |
Db 345 ACATCCTTGGATCGGACTAGCAACCAAGGACAGATATTCAGAGTGTACTCTAAA 404
| | | | |
QY 61 TCCAGTACCTCAAGCAAGTCCAGCAGCTAGCGTTGCCCAACTTAGATCAACAATGGTAG 120
| | | | |
Db 405 TCCAGTACCTCAAGCAAGTCCAGCAGCTAGCGTTGCCCAACTTAGATCAACAATGGTAG 464
| | | | |
QY 121 ACCGAGCATCAACTGTTTTCCTAAATAATGAAAGGTGAAGTGAACAGACTAAAGACA 180
| | | | |
Db 465 ACCGAGCATCAACTGTTTTCCTAAATAATGAAAGGTGAAGTGAACAGACTAAAGACA 524
| | | | |
QY 181 AACTGGAAACAGCCCAAAATGAATGAGTGGCTGGAAGTTTACGCCTGATAGCCAAACA 239
| | | | |
Db 525 AACTGGAAACAGCCCAAAATGAATGAGTGGCTGGAAGTTTACGCCTGATAGGTAAACA 583
| | | | |
RESULT 13
US-10-044-090-552
; Sequence 552, Application US/10044090
; Publication No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 552
; LENGTH: 1691
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 040057.1
US-10-044-090-552
Query Match 18.3%; Score 226.2; DB 13; Length 1691;
Best Local Similarity 96.7%; Pred. No. 2.9e-45;
Matches 231; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 1 ACATCCTTGGATCGGACTAGCAACCAAGGACAGATATTCAGAGTGTACTCTAAA 60
| | | | |
Db 409 ACATCCTTGGATCGGACTAGCAACCAAGGACAGATATTCAGAGTGTACTCTAAA 468
| | | | |
QY 61 TCCAGTACCTCAAGCAAGTCCAGCAGCTAGCGTTGCCCAACTTAGATCAACAATGGTAG 120
| | | | |
Db 469 TCCAGTACCTCAAGCAAGTCCAGCAGCTAGCGTTGCCCAACTTAGATCAACAATGGTAG 528
| | | | |
QY 121 ACCGAGCATCAACTGTTTTCCTAAATAATGAAAGGTGAAGTGAACAGACTAAAGACA 180
| | | | |
Db 529 ACCGAGCATCAACTGTTTTCCTAAATAATGAAAGGTGAAGTGAACAGACTAAAGACA 588
| | | | |

```
QY 181 AACTGGAACAGCCCAAAATGAAGTGGTCCCTGGGAAGTTTACGCTGTATAGCCAAACA 239
Db 589 AACTGGAACAGCCCAAAATGAAGTGGTCCCTGGGAAGTTTACGCTGTATAGCCAAACA 647

RESULT 14
US-10-044-090-551/c
; Sequence 551, Application US/10044090
; Publication No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0228 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 551
; LENGTH: 1996
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 040057.16
US-10-044-090-551

Query Match 18.3%; Score 226.2; DB 13; Length 1996;
Best Local Similarity 96.7%; Pred. No. 3.2e-45; Indels 0; Gaps 0;
Matches 231; Conservative 0; Mismatches 8;

QY 1 ACATCCTTGGGAATGGAGTACCAACCAAGGAGCAAGATATTCAAGAGTGTACTCTCAAA 60
Db 356 ACATCCTTGGTAATGGAGTACCAACCAAGGAGCAAGATATTCAAGAGTGTACTCTCAAA 297

QY 61 TCCAGTACCTCAAGCAAGTCCAGCAGCTAGCGTTGCCCACTTAGATCAACATGTGTAG 120
Db 296 TCCAGTACCTCAAGCAAGTCCAGCAGCGCGAGCGTTGCCCACTTAGATCAACATGTGTAG 237

QY 121 ACCAGCGCATCAACTTGTCTTAAATAATGAAGGTGAAGTGGAACTGGAACAGACTAAAGACA 180
Db 236 ACCAGCGCATCAACTTGTCTTAAATAATGAAGGTGAAGTGGAACTGGAACAGACTAAAGACA 177

QY 181 AACTGGAACAGCCCAAAATGAAGTGGTCCCTGGGAAGTTTACGCTGTATAGCCAAACA 239
Db 176 AACTGGAACAGCCCAAAATGAAGTGGTCCCTGGGAAGTTTACGCTGTATAGCCAAACA 118

RESULT 15
US-10-425-115-6885
; Sequence 6885, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 6885
; LENGTH: 694
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(694)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_106270C.1
```

US-10-425-115-6885

```
Query Match 16.8%; Score 207.6; DB 18; Length 694;
Best Local Similarity 91.6%; Pred. No. 7.4e-41; Indels 0; Gaps 0;
Matches 219; Conservative 0; Mismatches 20;

QY 1 ACATCCTTGGGAATGGAGTACCAACCAAGGAGCAAGATATTCAAGAGTGTACTCTCAAA 60
Db 383 ACATTCCTTGTCTATGCGGCTAGCAACCAAGAGCAGGAGATGCAAGAGTGCACCACTCAAA 442

QY 61 TCCAGTACCTCAAGCAAGTCCAGCAGCTAGCGTTGCCCACTTAGATCAACATGTGTAG 120
Db 443 TCCAGTACCTCAAGCAAGTTCAGCAGCGCGAGTGTGGCCCACTGAGATCAACAATGTGTAG 502

QY 121 ACCAGCGCATCAACTTGTCTTAAATAATGAAGGTGAAGTGGAACTGGAACAGACTAAAGACA 180
Db 503 ACCAGCGCATCAACTTGTCTTAAATAATGAAGGTGAAGTGGAACTGGAACAGACTAAAGACA 562

QY 181 AACTGGAACAGCCCAAAATGAAGTGGTCCCTGGGAAGTTTACGCTGTATAGCCAAACA 239
Db 563 NACTGGAAACAAGCCCCAAAATGAAGTGGTCCCTGGGAAGTTTACGCTGTATAGGTAAACA 621
```

Search completed: November 14, 2004, 16:21:12
Job time : 1977 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2004, 11:19:34 ; Search time 4402 seconds
(without alignments)
10248.157 Million cell updates/sec

Title: US-10-030-389-12
Perfect score: 1238
Sequence: 1 acatcttggaatgcagcta.....ttttgagggtccagttta 1238

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: gb_est1.*
2: gb_est2.*
3: gb_hic.*
4: gb_est3.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_gsi1.*
9: gb_gsi2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1014.4	81.9	1040	3	HSM802933
2	954.2	77.1	3765	3	CR627456 Homo sapi
3	785.6	63.5	1191	9	AY416358 Homo sapi
4	702.4	56.7	1191	9	AY416360 Mus muscu
5	692.8	56.0	860	5	BUI138450 603135077
6	622.8	50.3	1130	5	BX395380 BX395380
7	608.4	49.1	710	7	CN166400 997672 MA
8	604.4	48.8	758	4	B1089654 602854348
9	602.6	48.7	808	7	CO387170 AGNCOURT
10	595.6	48.1	710	5	BU942029 AGNCOURT
11	594.2	48.0	656	6	CB419819 592705 MA
12	590.6	47.7	702	4	B1185059 UNL-P-FN-
13	571.8	46.2	880	5	BU200958 603949240
14	566.4	45.8	660	1	AJ657839 AGNCOURT
15	553.4	44.7	961	6	CB182874 AGNCOURT
16	547.4	44.2	549	2	AW936487 QV4-DT002
17	545.6	44.1	883	5	BUI09878 603127357
18	543.6	43.9	925	1	AL895534 AGNCOURT
19	541.2	43.7	835	7	CO561762 AGNCOURT
20	537.8	43.4	757	7	CO389204 AGNCOURT
21	537.6	43.4	680	7	CK308195 SB0204582
22	536.2	43.3	884	5	BUI139100 603134268
23	529.6	42.8	544	2	AW753139 RCI-CT024
24	528.8	42.7	575	2	AW936415 QV4-DT002

25	518	41.8	947	5	BUI132234	BUI132234	603117225
26	510	41.2	606	4	BG896779	BG896779	HOA50-1-B
27	503.4	41.1	907	7	CF586305	CF586305	AGENCOURT
28	508.6	41.1	612	4	BG924897	BG924897	HNC38-1-E
29	505.8	40.9	862	5	BF165439	BF165439	BP165439
30	496	40.1	840	5	EX771427	EX771427	EX771427
31	492.2	39.8	805	5	BUI10001	BUI10001	603126168
32	489.8	39.6	617	5	BUI43216	BUI43216	603228940
33	489.6	39.5	637	1	AJ398971	AJ398971	AJ398971
34	485.4	39.2	700	1	AJ397697	AJ397697	AJ397697
35	477.2	38.5	799	5	BX665454	BX665454	BX665454
36	476	38.4	910	7	CF412751	CF412751	CH38081-G
37	471	38.0	548	6	CB287321	CB287321	CHM57-505
38	471	38.0	788	5	BUI44773	BUI44773	603125072
39	469	37.9	1161	5	BUI137716	BUI137716	603125072
40	461.2	37.3	667	6	CF254734	CF254734	mdvnl19-G
41	456.2	36.8	586	2	BF930054	BF930054	MR2-NT013
42	452.2	36.5	582	2	BF930039	BF930039	MR2-NT013
43	450.2	36.4	1031	5	BX387163	BX387163	BX387163
44	441.8	35.7	550	4	BGI42863	BGI42863	ia9e05-Y
45	441.2	35.6	814	6	CF258081	CF258081	phan105-I

ALIGNMENTS

RESULT 1
HSM802933
LOCUS Homo sapiens mRNA; cDNA DKFZp761K0722 (from clone DKFZp761K0722).
DEFINITION linear HTC 03-AUG-2004
ACCESSION AL583911
VERSION AL583911.1 GI:13093774
KEYWORDS HTC.
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1040)
Wambutt,R., Heubner,D., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
Fobo,G., Han,M. and Wiemann,S.
The German cDNA Consortium
Submitted (03-AUG-2004) MIPS, Ingolstaedter Landstr.1, D-85764
Neuherberg, GERMANY

CONSRFTM Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by Agowa (Berlin/Germany) within the cDNA sequencing
consortium of the German Genome Project.

TITLE This clone (DKFZp761K0722) is available at the RZPD Deutsches
consortium of the German Genome Project.
JOURNAL http://www.rzpd.de/cgi-bin/products/cl.cgi?cloneID=DKFZp761K0722
Further information about the clone and the sequencing project is
available at http://mips.gsf.de/projects/cdna/.

COMMENT

Location/Qualifiers
1..1040
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="RZPD:DKFZp761K0722"
/db_xref="taxon:9606"
/clone="DKFZp761K0722"
/tissue_type="amygdala"
/clone_lib="761 (synonym: hamy2). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev_stage="adult"
/note="Wilms' tumour 1-associating protein isoform 1,
N-terminus truncated"
1..1040
/gene="DKFZp761K0722"
3..860
/genes="DKFZp761K0722"
/codon_start=1
/product="hypothetical protein"

FEATURES

source
1..1040
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="RZPD:DKFZp761K0722"
/db_xref="taxon:9606"
/clone="DKFZp761K0722"
/tissue_type="amygdala"
/clone_lib="761 (synonym: hamy2). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev_stage="adult"
/note="Wilms' tumour 1-associating protein isoform 1,
N-terminus truncated"
1..1040
/gene="DKFZp761K0722"
3..860
/genes="DKFZp761K0722"
/codon_start=1
/product="hypothetical protein"

/protein_id="CAC29495.1"
 /db_xref="GI:13093775"
 /translation="MVDPAINLFLKMKLEQTKDLKLEQAQNELSAWKFPPDSQTKG
 KLMKCRMLIENBELGRQLSGRIQLEAEALQKKYSEBELKSSODELNDFFIQLDE
 EVEGMCSTILVLOOLKSTROQLAQYQQQSOASAPSTSRITASEPVEQSEATSKDCS
 RLTPNGSSSSSRORTSGSGFHRGNTTEDDPSSPGNGKSNSSSEPTGRGSGYV
 NOLSAYESVSPSPGSENLSHQNSDSDSHDPOBEKAVSGKRNITVGSRRVQNGLDS
 SVNQVGSVL"

ORIGIN

Query Match 81.9%; Score 1014.4; DB 3; Length 1040;
 Best Local Similarity 98.5%; Pred. No. 7.7e-228;
 Matches 1024; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

```

QY 112 CAATGGTAGACCCAGCGATCAACTTGTGTTTTCCTTAAATAAGAAAGGTGAACAGAGA 171
DB 1 CAATGGTAGACCCAGCGATCAACTTGTGTTTTCCTTAAATAAGAAAGGTGAACAGAGA 60
QY 172 TAAAGACAAACTTGAAACCAAGCCCAAAATGAAGTGGCTGGAAGTTTACGCCCTGATA 231
DB 61 CTAAGACAAACTTGAAACCAAGCCCAAAATGAAGTGGCTGGAAGTTTACGCCCTGATA 120
QY 232 GCCAAACAGGAAAGTAATGCGAAGTGTGCAATGCTTATCCAGGAGAAATCAAGAGC 291
DB 121 GCCAAACAGGAAAGTAATGCGAAGTGTGCAATGCTTATCCAGGAGAAATCAAGAGC 180
QY 292 TTGGAAGCGACGTGCCAGGAGCTATTGACCAACTTGAAGCAGAGTTCGCTTTACAGA 351
DB 181 TTGGAAGCGACGTGCCAGGAGCTATTGACCAACTTGAAGCAGAGTTCGCTTTACAGA 240
QY 352 AGAATAACAGTGAGGAGCTTAAAGCAGTCAGGATGAATGACTTCATCATCCAGC 411
DB 241 AGAATAACAGTGAGGAGCTTAAAGCAGTCAGGATGAATGACTTCATCATCCAGC 300
QY 412 TTGATGAGAGTAGAGGATGACAGATGACCATTCAGTTCCTGCAGCAGCAGCTGAGG 471
DB 301 TTGATGAGAGTAGAGGATGACAGATGACCATTCAGTTCCTGCAGCAGCAGCTGAGG 360
QY 472 AGACACGCCAGCAGTGTGGCTCAGTACCAGCAGCAGCAGTCTCAGGCTCTGCCCAAGTA 531
DB 361 AGACACGCCAGCAGTGTGGCTCAGTACCAGCAGCAGCAGTCTCAGGCTCTGCCCAAGTA 420
QY 532 CCAGCAGGACTACAGCTTCTGAACCTGTAGAAAGTCAGAGGCCAAGTAAGACTGCA 591
DB 421 CCAGCAGGACTACAGCTTCTGAACCTGTAGAAAGTCAGAGGCCAAGTAAGACTGCA 480
QY 592 GTCGCTCTGACAAACCGACCAAGTAATGCTGCTCCGCCAGAGAGAGCTCTGGCTCTG 651
DB 481 GTCGCTCTGACAAACCGACCAAGTAATGCTGCTCCGCCAGAGAGAGCTCTGGCTCTG 540
QY 652 GATTTCACAGGAGGGCAACACACCGAAGATGACTTCTCTTCTCCAGGGAATGGTA 711
DB 541 GATTTCACAGGAGGGCAACACACCGAAGATGACTTCTCTTCTCCAGGGAATGGTA 600
QY 712 ATAGTCTCTCAACAGCTCAGAGGAGAGACTGGCAGAGAGTAGTGGTTAGCTTAATC 771
DB 601 ATAGTCTCTCAACAGCTCAGAGGAGAGACTGGCAGAGAGTAGTGGTTAGCTTAATC 660
QY 772 AACTCAGTGGGGGTATGAAAGTGTAGACTCTCCACGGGCGAGTGAATACTCTCTCACAC 831
DB 661 AACTCAGTGGGGGTATGAAAGTGTAGACTCTCCACGGGCGAGTGAATACTCTCTCACAC 720
QY 832 ACCAATCAATGACACAGATCCAGTCATGACCTCAAGAGGAGAAAGCAGTAGTGGA 891
DB 721 ACCAATCAATGACACAGATCCAGTCATGACCTCAAGAGGAGAAAGCAGTAGTGGA 780
QY 892 AAGGTAATCGAACTGTGGGTTCGCCGCCACGTTTCAGAAATGCTTGGACTCAAGTGAATG 951
DB 781 AAGGTAATCGAACTGTGGGTTCGCCGCCACGTTTCAGAAATGCTTGGACTCAAGTGAATG 840
QY 952 TACAGGGTTCAGTTTGTGTAATATTTTTCAGGAAATTTTATACAGTGTCAATTTATTTG 1011
DB 841 TACAGGGTTCAGTTTGTGTAATATTTTTCAGGAAATTTTATACAGTGTCAATTTATTTG 900
  
```

QY 1012 GGAGAGGATACGTCTCAGAAAAATTAATGATCACTTTTGTCAAAATTCGCTTTTGTGGG 1071
 DB 901 GGAGAGGATACGTCTCAGAAAAATTAATGATCACTTTTGTCAAAATTCGCTTTTGTGGG 960
 QY 1072 TGTACGTTTGGTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 1131
 DB 961 TGTACGTTTGGTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 1020
 QY 1132 AAGAGGTGTTGTTGTTT 1151
 DB 1021 TTTTGTGTTTGTGTTT 1040

RESULT 2

CR627456 3765 bp mRNA linear HTC 03-AUG-2004
 LOCUS Homo sapiens mRNA; cDNA DKFP686F20131 (from clone DKFP686F20131).
 DEFINITION CR627456
 ACCESSION CR627456
 VERSION CR627456.1 GI:50949991
 KEYWORDS HTC.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 3765)
 AUTHORS Bloeker, H., Boecher, M., Brandt, P., Mewes, H.W., Weill, B., Amid, C.,
 Osanger, A., Fobo, G., Han, M. and Wiemann, S.
 CONSRTM The German cDNA Consortium
 TITLE Direct Submission
 JOURNAL Submitted (03-AUG-2004) MIPS, Ingolstaedter Landstr.1, D-85764
 Neuherberg, GERMANY
 COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
 sequenced by GBF (National Research Centre for Biotechnology Ltd.,
 Braunschweig/Germany) within the cDNA sequencing consortium of the
 German Genome Project.
 This clone (DKFP686F20131) is available at the RZPD Deutsches
 Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
 Please contact RZPD for ordering.
 http://www.rzpd.de/cgi-bin/products/cl.cgi?cloneID=DKFP686F20131
 Further information about the clone and the sequencing project is
 available at http://mips.gsf.de/projects/cdna/.

FEATURES

source
 1..3765
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="DKFP686F20131"
 /tissue_type="uterus"
 /clone_lib="f86 (synonym: hlcc3). Vector pSport1_sfi; host
 DH10B; sites SfiIA + SfiIB"
 /dev_stage="adult"
 /note="Wilm's tumour 1-associating protein isoform 1,
 N-terminus truncated, not fully spliced"
 1..3765
 /gene="DKFP686F20131"
 <2258..2298
 /gene="DKFP686F20131"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="CAH10537.1"
 /db_xref="GI:50949992"
 /translation="SQTGKLMKCRMLIENBELGRQLSGRIQLEAEALQKKYSEBELKSSODELNDFFIQLDE
 EELKSSODELNDFFIQLDEVEGMCSTILVLOOLKSTROQLAQYQQQSOASAPSTSRITASEPVEQSEATSKDCS
 RTTASEPVEQSEATSKDCSRLTPNGSSSRORTSGSGFHRGNTTEDDPSSPGNGKSNSSSEPTGRGSGYV
 NOLSAYESVSPSPGSENLSHQNSDSDSHDPOBEKAVSGKRNITVGSRRVQNGLDS
 SKGNRTVGSRRVQNGLDSNVQGSVL"

ORIGIN

Query Match 77.1%; Score 954.2; DB 3; Length 3765;
 Best Local Similarity 97.9%; Pred. No. 1.5e-213;
 Matches 986; Conservative 0; Mismatches 18; Indels 3; Gaps 2;

QY 231 AGCCAAACAGGGAAAAGTTAATGGCGAAGTGTGCAATGCTTATCCAGAGAAATCAAGAG 290
Db 2258 AGCCAAACAGGGAAAAGTTAATGGCGAAGTGTGCAATGCTTATCCAGAGAAATCAAGAG 2317
QY 291 CTTGGAAGGAGCTGTCCAGGAGCGTATTGCAACACTTGAAGCAGAGTTGGCTTTACAG 350
Db 2318 CTTGGAAGGAGCTGTCCAGGAGCGTATTGCAACACTTGAAGCAGAGTTGGCTTTACAG 2377
QY 351 AAGAAATACAGTGAAGGAGCTTAAAGACAGTCAGATGAATGAATGACTTCATCATCCAG 410
Db 2378 AAGAAATACAGTGAAGGAGCTTAAAGACAGTCAGATGAATGAATGACTTCATCATCCAG 2437
QY 411 CTTGATGAAGTGAAGGATGTCAGAGTACCATCTAGTTCTGAGCAGAGCTGAAG 470
Db 2438 CTTGATGAAGTGAAGGATGTCAGAGTACCATCTAGTTCTGAGCAGAGCTGAAG 2497
QY 471 GAGACAGCCAGAGTTGGTCTAGTACCAGCAGCAGCTCTCAGGCTCTGCCCCAAGT 530
Db 2498 GAGACAGCCAGAGTTGGTCTAGTACCAGCAGCAGCTCTCAGGCTCTGCCCCAAGT 2557
QY 531 ACCAGCAGGACTACAGTCTGCACTGTAGACAGTGTAGACAGTGTAGACAGTGTAG 590
Db 2558 ACCAGCAGGACTACAGTCTGCACTGTAGACAGTGTAGACAGTGTAGACAGTGTAG 2617
QY 591 AGTCGTCTGCAACAGGACCAAGTAATGGTAGCTCTCCGCGCAGAGGAGCTGTGGGTCT 650
Db 2618 AGTCGTCTGCAACAGGACCAAGTAATGGTAGCTCTCCGCGCAGAGGAGCTGTGGGTCT 2677
QY 651 GGATTTACAGGGAGGAGCAACACACCGAAGATGACTTCTCTCTCTCAGGGAATGGT 710
Db 2678 GGATTTACAGGGAGGAGCAACACACCGAAGATGACTTCTCTCTCTCAGGGAATGGT 2737
QY 711 AATAGTCTCCACAGCTCAGAGGAGAGAACTGGCAGAGAGTGTAGTGTACGTAAAT 770
Db 2738 AATAGTCTCCACAGCTCAGAGGAGAGAACTGGCAGAGAGTGTAGTGTACGTAAAT 2797
QY 771 CAATCAGTGTGGGATGAAAGTGTAGACTCTCCACGGGAGTGAATACTCTCTACA 830
Db 2798 CAATCAGTGTGGGATGAAAGTGTAGACTCTCCACGGGAGTGAATACTCTCTACA 2857
QY 831 CACCAATCAATGACAGACTCAGTCAATGACCTCAAGAGGAGAGAGTGTAGTGGG 890
Db 2858 CACCAATCAATGACAGACTCAGTCAATGACCTCAAGAGGAGAGAGTGTAGTGGG 2917
QY 891 AAGGTAAATCGAACTGTGGGTTCCGCGCAGCTCAGAAATGGCTTGGACTCAAGTAAAT 950
Db 2918 AAGGTAAATCGAACTGTGGGTTCCGCGCAGCTCAGAAATGGCTTGGACTCAAGTAAAT 2977
QY 951 GTACAGGTTGAGTTTGTATATTTTTCAGGAAATTTTATACAGTGTCAATTAATTT 1010
Db 2978 GTACAGGTTGAGTTTGTATATTTTTCAGGAAATTTTATACAGTGTCAATTAATTT 3037
QY 1011 GGGAGAGGATCTGTCCAGAAAATTAATGATATCTTTTGTGTCACAAATTTGCTTTTGTGG 1070
Db 3038 GGGAGAGGATCTGTCCAGAAAATTAATGATATCTTTTGTGTCACAAATTTGCTTTTGTGG 3097
QY 1071 GTGACGTTTGTGTTTTTTTGTGTTTTTTTCTTTGGAATTTTTTTCTTTCTTTT 1130
Db 3098 GTGACGTTTGTGTTTTTTTGTGTTTTTTTCTTTGGAATTTTTTTCTTTCTTTT 3157
QY 1131 TAAGAGTGTGTTGTTTTTTTGTGTTTTTTTCTTTGGAATTTTGTGGAATTTGTAACAGT 1190
Db 3158 T--TTTTTTTTTTTTTTTTTTTGTGTTTTTTTCTTTGGAATTTTGTGGAATTTGTAACAGT 3215
QY 1191 TAAATTAATTTGAATTTGC-AAAAGGGCAATTTTGTGAGGCTTCAGTTA 1238
Db 3216 TAAATTAATTTGAATTTGCTTAAAGGACATTTTGTGAGGCTCAAGTTA 3264

RESULT 3
AY416358
LOCUS

AY416358 1191 bp DNA linear GSS 12-DEC-2003

DEFINITION Homo sapiens HCM5864 gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION AY416358
VERSION AY416358.1 GI:39772318
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1191)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 1191)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source
1..1191
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
<1..>1191
/locus_tag="HCM5864"
ORIGIN
Query Match 63.5%; Score 785.6; DB 9; Length 1191;
Best Local Similarity 81.2%; Pred. No. 5.9e-174;
Matches 783; Conservative 0; Mismatches 183; Indels 0; Gaps 0;
QY 1 ACATCCTTGGATGCGACTAGCAACCAAGGAGCAAGATATTCAAGAGTGTACTCTCAA 60
Db 221 ACATCCTTGGATGCGACTAGCAACCAAGGAGCAAGATATTCAAGAGTGTACTCTCAA 280
QY 61 TCCAGTACTCAAGCAAGTCCAGCAGCTAGGCTGCCCAACTTAGATCAACAATGGTAG 120
Db 281 NNN 340
QY 121 ACCAGCGATCACTTGTTTTCCTTAAATAATGAAAGGTGAAGTGAAGTGAAGTGAAG 180
Db 341 NNN 400
QY 181 AACTGGAAACAAGCCCAAAATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 240
Db 401 NNN 460
QY 241 GGAAGAAAGTAAAGGAGTGTGGAAGTGTGGAAGTGTGGAAGTGTGGAAGTGTGGAAGTGTGGAAG 300
Db 461 GGAAGAAAGTAAAGGAGTGTGGAAGTGTGGAAGTGTGGAAGTGTGGAAGTGTGGAAGTGTGGAAG 520
QY 301 AGCTGTCCAGGAGAGTATTGCAACAATTTGAAGCAGAGTGTGCTTTACAGAGAAATACA 360
Db 521 AGCTGTCCAGGAGAGTATTGCAACAATTTGAAGCAGAGTGTGCTTTACAGAGAAATACA 580
QY 361 GTGAGAGCTTAAAGCAGTCAAGTGAATGACTTCAATCATCCAGCTTCAAGCTTGAAG 420
Db 581 GTGAGAGCTTAAAGCAGTCAAGTGAATGACTTCAATCATCCAGCTTCAAGCTTGAAG 640
QY 421 AAGTAGAGGTATGACAGAGTACCATTTCTAGTTCTGAGCAGCAGCTGTGAAGGACACGCC 480
Db 641 AAGTAGAGGTATGACAGAGTACCATTTCTAGTTCTGAGCAGCAGCTGTGAAGGACACGCC 700

```

QY 481 AGCAGTTGGCTCAGTACCAGCAGCAGAGTCTCAGGCTCTGCCCAAGTACCAGCAGGA 540
Db 701 AGCAGTTGGCTCAGTACCAGCAGCAGCAGTCTCAGGCTCTGCCCAAGTACCAGCAGGA 760

QY 541 CTACAGCTTCTGAACCTCTAGAACAGTCTCAGAGGCCACAAAGTAAAGACTGCGAGTCTGTA 600
Db 761 CTACAGCTTCTGAACCTCTAGAACAGTCTCAGAGGCCACAAAGTAAAGACTGCGAGTCTGTA 820

QY 601 CAACGGACCAAGTAAGTCTCTCTCCCGCCAGAGAGAGTCTGGGTCTGGATTTTACA 660
Db 821 CAACGGACCAAGTAAGTCTCTCTCCCGCCAGAGAGAGTCTGGGTCTGGATTTTACA 880

QY 661 GGGAGGGCAACACAAACCAAGATGACTTCTCTCTCCAGGGAATGGTAATAAGTCCT 720
Db 881 GGGAGGGCAACACAAACCAAGATGACTTCTCTCTCCAGGGAATGGTAATAAGTCCT 940

QY 721 CCAACAGCTCAGAGAGAGAACTGGCAGAGAGAGTGTAGTAAATCAACTCAGTG 780
Db 941 CCAACAGCTCAGAGAGAGAACTGGCAGAGAGAGTGTAGTAAATCAACTCAGTG 1000

QY 781 CGGGTATGAAGTGTAGACTCTCCACGGGCAAGTCTCTCACACCAATCAA 840
Db 1001 CGGGTATGAAGTGTAGACTCTCCACGGGCAAGTCTCTCACACCAATCAA 1060

QY 841 ATGACACAGACTCCAGTCAATGACCTCAAGAGAGAGAGTGTGGGAAAGGTAATC 900
Db 1061 ATGACACAGACTCCAGTCAATGACCTCAAGAGAGAGAGTGTGGGAAAGGTAATC 1120

QY 901 GAATGTGGGTCCCGCCAGCTTCAGATGGCTTGGACTCAAGTGTAAATGTACAGGGT 960
Db 1121 GAATGTGGGTCCCGCCAGCTTCAGATGGCTTGGACTCAAGTGTAAATGTACAGGGT 1180

QY 961 CAGTTTGTAA 971
Db 1181 CAGTTTGTAA 1191

RESULT 4
AY416360
LOCUS
DEFINITION Mus musculus HCM5864 gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION AY416360
VERSION AY416360.1 GI:39772320
SOURCE GSS.
ORGANISM Mus musculus (house mouse)
REFERENCE
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source
1..1191
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"

```

```

<1..1191
/locus_tag="HCM5864"

Gene
ORIGIN
Query Match 56.7%; Score 702.4; DB 9; Length 1191;
Best Local Similarity 75.6%; Pred No 2,3e-154;
Matches 736; Conservative 0; Mismatches 233; Indels 0; Gaps 0;

QY 1 ACATCTTGGAATGCGACTAGCAACCAAGGACAAAGATATTCAAGAGTGTACTCTCAA 60
Db 221 ACATCTTGGAATGCGACTAGCAACCAAGGACAAAGATATTCAAGAGTGTACTCTCAA 280

QY 61 TCCAGTACTCTCAAGCAAGTCTCAGCAGCCCTAGGTTGCCCACTTAGATCAACAATGGTAG 120
Db 281 TCCAGTACTCTCAAGCAAGTCTCAGCAGCCCTAGGTTGCCCACTTAGATCAACAATGGTAG 340

QY 121 ACCCAGCGATCAACTTGTCTTCTTAAATAATGAAAGGTGAACTGGACAGACTAAAGACA 180
Db 341 ACCCAGCGATCAACTTGTCTTCTTAAATAATGAAAGGTGAACTGGACAGACTAAAGACA 400

QY 181 AACTGGAAACAGCCCAAAATGAACTGAGTGCCTGGAAGTTTACGCCTGATAGCCAAACAG 240
Db 401 AACTGGAAACAGCCCAAAATGAACTGAGTGCCTGGAAGTTTACGCCTGATAGCCAAACAG 460

QY 241 GGAATAAGTAAATGGCGAAGTGTGCAATGCTTATCCAGGAGAACTCAAGAGCTTGGAAAGGC 300
Db 461 GGAATAAGTAAATGGCGAAGTGTGCAATGCTTATCCAGGAGAACTCAAGAGCTTGGAAAGGC 520

QY 301 AGCTGTCCAGGAGAGTATTCGACAACTTGAAGCAGAGTGTGGCTTTTACAGAGAAATACA 360
Db 521 AGCTGTCCAGGAGAGTATTCGACAACTTGAAGCAGAGTGTGGCTTTTACAGAGAAATACA 580

QY 361 GTGAGGAGCTTAAAGAGCAGTCAAGATGAACTGATCTTCAATCATCCAGCTTCAATCAAG 420
Db 581 GTGAGGAGCTTAAAGAGCAGTCAAGATGAACTGATCTTCAATCATCCAGCTTCAATCAAG 640

QY 421 AAGTAGGGTATGAGAGTACCACTTCTAGTCTTGCAGCAGCAGCTGAAAGAGACAGCC 480
Db 641 AAGTAGGGTATGAGAGTACCACTTCTAGTCTTGCAGCAGCAGCTGAAAGAGACAGCC 700

QY 481 AGCAGTTGGCTCAGTACCAGCAGCAGTCTCAGGCTCTGCCCAAGTACCAGCAGGA 540
Db 701 AGCAGTTGGCTCAGTACCAGCAGCAGTCTCAGGCTCTGCCCAAGTACCAGCAGGA 760

QY 541 CTACAGCTTCTGAACCTCTAGAACAGTCTCAGAGGCCACAAAGTAAAGACTGCGAGTCTGTA 600
Db 761 CTACAGCTTCTGAACCTCTAGAACAGTCTCAGAGGCCACAAAGTAAAGACTGCGAGTCTGTA 820

QY 601 CAACGGACCAAGTAAAGTCTCTCCCGCCAGAGAGAGTCTGGGTCTGGATTTTACA 660
Db 821 CAACGGACCAAGTAAAGTCTCTCCCGCCAGAGAGAGTCTGGGTCTGGATTTTACA 880

QY 661 GGGAGGGCAACACAAACCAAGATGACTTCTCTCTCCAGGGAATGGTAATAAGTCCT 720
Db 881 GGGAGGGCAACACAAACCAAGATGACTTCTCTCTCCAGGGAATGGTAATAAGTCCT 940

QY 721 CCAACAGCTCAGAGAGAGAACTGGCAGAGAGAGTGTAGTAAATCAACTCAGTG 780
Db 941 CCAACAGCTCAGAGAGAGAACTGGCAGAGAGAGTGTAGTAAATCAACTCAGTG 1000

QY 781 CGGGTATGAAGTGTAGACTCTCCACGGGCAAGTCTCTCACACCAATCAA 840
Db 1001 CGGGTATGAAGTGTAGACTCTCCACGGGCAAGTCTCTCACACCAATCAA 1060

QY 841 ATGACACAGACTCCAGTCAATGACCTCAAGAGAGAGAGTGTGGGAAAGGTAATC 900
Db 1061 ATGACACAGACTCCAGTCAATGACCTCAAGAGAGAGAGTGTGGGAAAGGTAATC 1120

QY 901 GAATGTGGGTCCCGCCAGCTTCAGATGGCTTGGACTCAAGTGTAAATGTACAGGGT 960
Db 1121 GAATGTGGGTCCCGCCAGCTTCAGATGGCTTGGACTCAAGTGTAAATGTACAGGGT 1180

QY 961 CAGTTTGTAA 971

```

```

Db      1181  CAGTTTGTAA 1191
|||||
RESULT 5
BUI38450      860 bp  mRNA  linear  EST 25-NOV-2002
LOCUS        603135077F1 CSEQHL24 Gallus gallus cDNA clone CHEST117j19 5', mRNA
DEFINITION
ACCESSION    BUI38450.1 GI:25352616
VERSION      BUI38450
KEYWORDS     EST.
SOURCE       Gallus gallus (chicken)
ORGANISM     Gallus gallus
REFERENCE    1 (bases 1 to 860)
AUTHORS     Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Eurt,D.W., Bosch,E.,
Fong,W.T., Rickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
TITLE       A Comprehensive Collection of Chicken cDNAs
JOURNAL     Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE     22335534
PUBMED      12445392
COMMENT     Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology
(UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
FEATURES
source
1..860
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, Hisex"
/db_xref="taxon:9031"
/clone="CHEST117j19"
/dev_stage="16 day embryo"
/lab_host="DRI08"
/clone_lib="CSEQCHL24"
/note="Organ: brain; Vector: pBluescript II KS(+); Site: 1:
EcoRI; Site 2: NotI; Modification of pBluescript II KS(+):
[Stratagene] vector to accommodate cDNA produced with the
T-trimmed protocol (Construction of uni-directionally
cloned cDNA libraries from messenger RNA for improved 3'
end DNA sequencing by Glenn Fu, et al. U.S. Patent #
6,387,624). Cut pBluescript II KS(+) with NotI and EcoRI.
Ligate in double stranded adaptor containing BspI and
BamHI sites [5'ggccgcgtgcagcccgatccgaaaaag]
[5'aattcttttttcggatccgggtgcagc]"
ORIGIN
Query Match 56.0%; Score 692.8; DB 5; Length 860;
Best Local Similarity 89.6%; Pred. No. 3.9e-152;
Matches 756; Conservative 0; Mismatches 87; Indels 1; Gaps 1;
Qy      81  CAGCAGCTAGCGTTGCCCACTTAGATCAACAATGGTAGACCAGCGATCAACTGTTT 140
Db      10  CAGCAGCTAGCGTTGCCCACTTAGATCAACAATGGTAGACCAGCGATCAACTGTTT 69
Qy      141 TTCCTAAATGAAAGTGAATCGAAGACAGACTAAGACAACTGGAACAGCCCAAAAT 200
Db      70  TTCCTAAATGAAAGTGAATCGAAGACAGACTAAGACAACTGGAACAGCCCAAAAT 129
Qy      201 GAACCTAGCTGCTGGAAGTTTACCCCTGATAGCCCAACAGAGGAAAAGTTAATGCGGAAG 260
Db      130 GAACCTAGCTGCTGGAAGTTTACCCCTGATAGCCCAACAGAGGAAAAGTTAATGCGGAAG 189
Qy      261 TGTCGAATGCTTATCCAGAGAAATCAAGAGCTTGGAAAGCGCAGCTGTCCAGGGACGTATT 320
Db      190 TGTCGAATGCTTATCCAGAGAAATCAAGAGCTTGGAAAGCGCAGCTGTCCAGGGACGTATT 249

```

```

Qy      321 GCACAACCTGAAGCAGAGTTGGCTTTTACAGAGAAATACAGTGAAGGAGCTTAAAGCAGT 380
Db      250 GCACAGCTTGAGGCAGAGTTGGCTTTTACAGAGAAATATAGTGAAGAACCTAAAGCAGT 309
Qy      381 CAGATGAACCTGAATGACTTTCATCATCCAGCTTGATGAAGAAAGTAGAGGGTATCCAGAGT 440
Db      310 CAGATGAATGATGACTTTCATCATCCAGCTTGATGAAGAAAGTAGAGGGTATCCAGAGT 369
Qy      441 ACCATTCTAGTTCTGACGACGACGCTGAAGAGACACGCCAGCAGTTGGCTCAGTACCAG 500
Db      370 ACCATTCTAGTTCTGACGACGACGCTGAAGAGAGACTCGCCAGCAGTTGGCCGACGATCCAG 429
Qy      501 CAGCAGCAGTCTCAGGCTCTGCCCAAGTACCAGCAGGACTACAGCTTCTGAACTCTGTA 560
Db      430 CAGCAGCAGTCTCAGGCTCTGCCCAAGTACCAGCAGGACTTCCATCTTCTGAGCCTTACA 489
Qy      561 GAACAGCTCAGAGGGCCACAAGTAAGACTGCGTCTGTGACAAACGACCAAGTAATGGT 620
Db      490 GACCAG-CAGAGGGCTGTGGGTAAGACTGTCAGCCGCTCTGGCTAACCGACCAAGCAATGGT 548
Qy      621 AGCTCTCTCCGCGCAGAGGACGCTCTGGGCTCTGGATTTACAGAGGGGGGCAACACACCGAA 680
Db      549 AGCTCTCTCCGCGCAGAGGACGCTCTGGGCTCTGGATTTATAGGGAGGGTAGCAGCAGAA 608
Qy      681 GATGACTTTTCTTCTTCTCCAGGGAATGCTAATAGTCTCTCCAAACAGCTCAGAGGAGAGA 740
Db      609 GATGACTTTCCCGGCTTCTCCAGGGAATGCTAATAGTCTCCAAACAGCTCTGAAGATAGA 668
Qy      741 ACTGCAGAGAGGAGTAGTGGTTACGTAATCAACTCAGTGGGGGTATGAAGTGTAGAC 800
Db      669 ACTGGTAGAGAGGAGTGGTAGCTACATAAACCAACTCAGTACTGGGTATGAAGTGTAGAC 728
Qy      801 TCTCCACGGGCGAGTGAAGAACTCTCTCACACACCAATCAATGACACAGACTCCAGTCAAT 860
Db      729 TCTCCACGGGCGAGTGAAGAACTCTCTCACACCAATCAATGACACAGACTCCAGTCAAT 788
Qy      861 GACCTCTAAGAGGAGAGGAGTGGAGTGGGAAAGTAATCGAATCTGTGGGTTCCCGCCAC 920
Db      789 GATCTCTAAGAGGAGAGGAGTGAAGTGAAGGTAACAGAACTGCGGGTTCTCGGCAT 848
Qy      921 GTTC 924
Db      849 GTCC 852
RESULT 6
BX395380      1130 bp  mRNA  linear  EST 01-MAY-2004
LOCUS        BX395380 Homo sapiens NEUROBLASTOMA COT 50-NORMALIZED Homo sapiens
DEFINITION   cDNA clone CS0DD006YF24 5-PRIME, mRNA sequence.
ACCESSION    BX395380
VERSION      BX395380.2 GI:46920201
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 1130)
AUTHORS     Li,W.B., Gruber,C., Jesse,J. and Polayes,D.
TITLE       Full-length cDNA libraries and normalization
JOURNAL     Unpublished (2001)
COMMENT     On May 13, 2003 this sequence version replaced gi:30628516.
Contact: Genoscope
Genoscope - Centre National de Sequencage
Bp 191 91006 EVRY cedex - France
Email: seqrefgenoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies a
division of Invitrogen. This sequence belongs to sequence cluster
3991.f

```

For more information about this cluster, see
<http://www.genoscope.cns.fr/cdna?c=CS0DD006DHL2QPl&c=3991.f>.

FEATURES

Location/Qualifiers
 1..1130
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DD006VP24"
 /tissue type="NEUROBLASTOMA COT 50-NORMALIZED"
 /clone_lib="Homo sapiens NEUROBLASTOMA COT 50-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
 Query Match 50.3%; Score 622.8; DB 5; Length 1130;
 Best Local Similarity 90.8%; Pred. No. 1.2e-135;
 Matches 652; Conservative 26; Mismatches 38; Indels 2; Gaps 2;
 QY 1 ACATCCCTTGGATGCGACTAGCAACCAAGGACAGATATTCAGAGTGTACTACTCAA 60
 DB 414 ACATCCCTTGGATGCGACTAGCAACCAAGGACAGATGCGAGTGTACTACTCAA 473
 QY 61 TCCAGTACCTCAAGCAAGTTCAGAGCGCGGTGCGCAACTGAGATCAACAATGTAG 120
 DB 474 TCCAGTACCTCAAGCAAGTTCAGAGCGCGGTGCGCAACTGAGATCAACAATGTAG 533
 QY 121 ACCCAGCGATCAACTGTTTCTTAAATGAAAGTGAAGTGGACAGACTAAAGACA 180
 DB 534 ACCCAGCGATCAACTGTTTCTTAAATGAAAGTGAAGTGGACAGACTAAAGACA 593
 QY 181 AACTGGAAACAGCCAAATGAATGAGTGCCTGGAAAGTTTACGCTGTAGCCAAACAG 240
 DB 594 AACTGGAAACAGCCAAATGAATGAGTGCCTGGAAAGTTTACGCTGTAGCCAAACAG 653
 QY 241 GGAATAAGTTAATGGCAAGTGTGCAATGCTTATCCAGGAGATCAAGAGCTTGGAGGC 300
 DB 654 GGAATAAGTTAATGGCAAGTGTGCAATGCTTATCCAGGAGATCAAGAGCTTGGAGGC 713
 QY 301 AGCTGTCCAGGGACGATTGCAAACTTGAAGCAGAGTTGGCTTTACAGAGAAATACA 360
 DB 714 AGCCGTRCCAGGGACGATTGCAAACTTGAAGCAGAGTTGGCTTTACAGAGAAATACA 773
 QY 361 GTGAGGAGCTTAAAGCAGTCAGATGAATGATGACTTCACTATCCAGCTTGTATGAG 420
 DB 774 GTWAGGAGCTTAAAGCAGTCAGATGAATGATGACTTCACTATCCAGCTTGTATGAG 833
 QY 421 AAGTAGAGGTTATGAGAGTACCATTTCTAGTTCTGAGCAGCAGCTG-AAGGAGACAGC 479
 DB 834 AAGTAGAGGTTATGAGAGTACCATTTCTAGTTCTGAGCAGCAGCTG-AAGGAGAAAGC 893
 QY 480 CAGCAGTTGCTGATGACAGCAGCAGCTCTCAGGCTCTGCCCAAGTACCAGCAG 539
 DB 894 CAGCAGTTGCTGATGACAGCAGCAGCTCTCAGGCTCTGCCCAAGT-MCAACMG 952
 QY 540 ACTACAGCTTCTGAACCTGTAGAACAGTCAGAGCCACAGTAAGACTGCACTGCTG 599
 DB 953 ACTACAGCTTCTGAACCTGTAGAACAGTCAGAGCCCAAGTAAGACTGCMGTGTTG 1012
 QY 600 ACAAACGCAAGTAATGTAGTCTCTCCCGCAGAGCAGCTCTGGGTCTGGATTTCAC 659
 DB 1013 ASAAACGCAAGTAATGTAGTCTCTCCCGCAGAGCAGCTCTGGGTCTGGATTTCAC 1072
 QY 660 AGGAGGGCAACCAACCGAAGTACTTTCCTTCTCCAGGGGAATGTAATAGT 717
 DB 1073 AGGAGGGCAACCAACCGAAGTACTTTCCTTCTCCAGGGGAATGTAATAGT 1130

RESULT 7

CN166400/c 710 bp mRNA linear EST 02-APR-2004
 LOCUS
 DEFINITION 997672 MARC 4P1G Sus scrofa cDNA 3', mRNA sequence.

ACCESSION

CN166400.1 GI:46180830

VERSION

EST.

KEYWORDS

Sus scrofa (pig)

SOURCE

Sus scrofa

ORGANISM

Sus scrofa

REFERENCE

1 (bases 1 to 710)

AUTHORS

Smith, T.P.L., Fraking, B.A., Ford, J.J., Vallet, J.L., Wise, T.A.,

TITLE

Noneman, D.J., Wray, J.E. and Keele, J.W.

JOURNAL

Porcine EST collection using a normalized library constructed from

COMMENT

embryos representing early developmental stages

Unpublished (2003)

Contact: Smith TPL

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called with phred v0.020425.c and

trimmed with the aid of the trim_alt option. Vector identified with

cross_match v0.990329.

Plate: TM8066 row: L column: 12

Seq primer: TAGAAGGCACAGTCGAGG.

Location/Qualifiers

1..710

/organism="Sus scrofa"

/mol_type="mRNA"

/db_xref="taxon:9823"

/tissue_type="pooled"

/lab_host="DH10B"

/clone_lib="MARC 4P1G"

/note="Vector: pCDNA3.1; Site 1: EcoRI; Site 2: NotI;

Library made with combined RNA from day-10, day-13,

day-15, day-25, and day-30 whole embryos."

ORIGIN

Query Match 49.1%; Score 608; DB 7; Length 710;
 Best Local Similarity 91.9%; Pred. No. 3.4e-132;
 Matches 654; Conservative 0; Mismatches 55; Indels 3; Gaps 1;
 QY 262 GTGCAATGCTTATCCAGGAGATCAAGAGCTTGAAGGAGCTGTCGCCAGGAGCTATTG 321
 DB 710 GTGCAATGCTTATCCAGGAGATCAAGAGCTTGAAGGAGCTGTCGCCAGGAGCTATTG 651
 QY 322 CACAACCTTGAAGCAGAGTGTGGCTTTTACAGAGAAATACAGTGAGGAGCTTAAAGCAGTC 381
 DB 650 CACAACCTTGAAGCAGAGTGTGGCTTTTACAGAGAAATACAGTGAGGAGCTTAAAGCAGTC 591
 QY 382 AGGATGAGCTTGAAGTGTGCTTATCCAGCTTGTAGAGAGTAGAGGCTATCCAGAGTA 441
 DB 590 AGGATGAGCTTGAAGTGTGCTTATCCAGCTTGTAGAGAGTAGAGGCTATCCAGAGTA 531
 QY 442 CCATTCTAGTTCAGCAGCAGCTGAAGGAGACACCCAGCAGCTTGGCTCAGTACCAGC 501
 DB 530 CCATTCTAGTTCAGCAGCAGCTGAAGGAGACCCCCAGCAGCTTGGCTCAGTACCAGC 471
 QY 502 AGGAGCAGCTCAGGCTCTGCCCAAGTACCAGCAGCTACAGCTTCTGACCTGTAG 561
 DB 470 AGGAGCAGCTCAGGCTCTGCCCAAGTACCAGCAGCTACAGCTTCTGACCTGTAG 414
 QY 562 AACAGTCAGAGGCGCACAAAGTAAAGCTGCACTGCTGTGACAAACGAGACCAAGTAATGGTA 621
 DB 413 GACAGGAGAGGCGCACAAAGTAAAGCTGCACTGCTGTGACCAACGAGACCAAGTAATGGTA 354
 QY 622 GCTCCTCCCGCAGAGAGCTCTGGGTCTGGATTTTACAGGGAGGGCAACACAAACCGAAG 681
 DB 353 GCTCCTCCCGCAGAGAGCTCTGGGTCTGGATTTTACAGGGAGGGCAACACAAACCGAAG 294
 QY 682 ATGACTTCTCTTCTTCCAGGGAATGTAATAGTCTCTCAACAGCTCAGAGGAGAGAA 741
 DB 293 ATGACTTCTCTTCTTCCAGGGAATGTAATAGTCTCTCAACAGCTCAGAGGAGAGAA 234


```

QY 742 CTGGCAGAGAGTAGTGGTTAGCTAAATCAACTCAGTCGGGGTATGAAGTGTAGACT 801
Db 233 CTGGCAGAGAGTAGTGGTTAGCTAAATCAACTCAGTCGGGGTATGAAGCGTAGACT 174
QY 802 CTCCACGGGCACTGAAACTCTCTCACACACCAATCAAAATGACACAGACTCCAGTCATG 861
Db 173 CTCCACGGGCACTGAAACTCTCTCACACGGCACTCACATGACACAGACTCCAACTG 114
QY 862 ACCTCAAGAGGAGAAAGCAGTGGAGTGGAAAGGTAATGAACTGCGGTTCGCCGACG 921
Db 113 ACCTCAAGAGGAGAAAGCAGTGGAGTGGAAAGGTAATGAACTGCGGTTCGCCGACG 54
QY 922 TCCAGATGCTTGGACTCAAGTGTAAATGTACAGGGTTCAGTTTGTAAATA 973
Db 53 TCCAGATGCTTGGACTCAAGTGTAAATGTACAGGGTTCGGTTTGTAAAA 2

RESULT 8
BI089654
LOCUS 602854348F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:4996066 5',
DEFINITION mRNA sequence.
ACCESSION BI089654
VERSION 1.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 758)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHAM1020 row: 0 column: 11
High quality sequence start: 8
High quality sequence stop: 622.
FEATURES
source
1..758
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4996066"
/cell_line="MGC36"
/lab_host="DH10B"
/clone_lib="NIH_MGC_10"
/notes="Organ: cervix; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: oligo dr.
Average insert size 1.5 kb. Library prepared by Life
Technologies."

ORIGIN
Query Match 48.8%; Score 604.4; DB 4; Length 758;
Best Local Similarity 92.6%; Pred. No. 2.5e-131;
Matches 646; Conservative 0; Mismatches 51; Indels 1; Gaps 1;

QY 476 ACGCCAGAGTGGCTCAGTACACGACGACGAGCTTCAGCGCTCTCCGCCAAGTACCAG 535
Db 13 ACGCCAGAGTGGCTCAGTACACGACGACGAGCTTCAGCGCTCTCCGCCAAGTACCAG 72
QY 536 CAGGACTACAGCTTCTGACCTGTAGAACAGTACAGGCCACAGTAAGACTGCAGTCG 595
Db 73 CAGGACTACAGCTTCTGACCTGTAGAACAGTACAGGCCACAGTAAGACTGCAGTCG 132
QY 596 TCTGACAAACGACCAAGTAATGGTAGCTCTCCGCCGACGAGGACGCTCTGGGTCTGGATT 655

```

```

133 TCTGACAAACGGACCAAGTAATGGTAGCTCTCCGCCGACGAGGACGCTCTGGGTCTGGATT 192
QY 656 TCACAGGAGGAGGCAACCAACACCGAAGATGATCTTCTTCTTCTCAGGGATGATATAAA 715
193 TCACAGGAGGAGGCAACCAACACCGAAGATGATCTTCTTCTTCTCAGGGATGATATAAA 252
QY 716 GTCTTCCAAACAGCTCAGAGGAGAGAACTGGCAGAGGAGGTAGTGGTTACGTAATAACT 775
253 GTCTTCCAAACAGCTCAGAGGAGAGAACTGGCAGAGGAGGTAGTGGTTACGTAATAACT 312
QY 776 CAGTGGCGGGTATGAAAGTGTAGACTCTCCACACGGGAGTGAATACTCTCTCACACACCA 835
313 CAGTGGCGGGTATGAAAGTGTAGACTCTCTCCACACGGGAGTGAATACTCTCTCACACACCA 372
QY 836 ATCAAAATCACACAGACTCCAGTCATGACCCCTCAAGAGGAGAGAAAGCAGTGGAGGAAGG 895
373 ATCAAAATCACACAGACTCCAGTCATGACCCCTCAAGAGGAGAGAAAGCAGTGGAGGAAGG 432
QY 896 TAATCGAACTGTGGGTTCGCCGCCACGTTCAAGATGGCTTGGACTCAAGTGTAAATGTACA 955
433 TAATCGAACTGTGGGTTCGCCGCCACGTTCAAGATGGCTTGGACTCAAGTGTAAATGTACA 492
QY 956 GGGTTCAGTTTGTAAATA-TTTTTTCAGCAAAATTTTATACAGTGTCAATTAATTCGGA 1014
493 GGGTTCAGTTTGTAAATA-TTTTTTCAGCAAAATTTTATACAGTGTCAATTAATTCGGA 552
QY 1015 CAGGATACGTCCAGAAAATTAATGCATACATTTTGTACAAATTTGCGCTTTTGGGTGT 1074
553 CAGGATACGTCCAGAAAATTAATGCATACATTTTGTACAAATTTGCGCTTTTGGGTGT 612
QY 1075 ACGTTTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1134
613 ACGTTTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 672
QY 1135 AGGTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1172
673 AGGTAAATAAATAAAGAGGGGCAAAATTTGGGACG 710

CO387170 808 bp mRNA linear EST 30-JUN-2004
AGENCOURT 26625081 NIH_MGC_253 Rattus norvegicus cDNA clone
IMAGE:7302938 5', mRNA sequence.
CO387170
CO387170.1 GI:49494761
EST.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 808)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Drs. Josef Lazar & Howard Jacob, Medical
College of Wisconsin
cDNA Library Preparation: Express Genomics
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHAM15337 row: b column: 24
High quality sequence stop: 654.
Location/Qualifiers
FEATURES

```



```

source
1. 808
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="IMAGE:7302938"
/tissue type="Ovary - Brown Norway Line 3 Age 8 weeks,
tissues were snap-frozen and transferred in -70C. RNase
free the entire procedure."
/lab host="DH10B Tena"
/clone_lib="NIH_MGC_253"
/notes="Organ: ovary; Vector: pExpress-1; Site:1: EcoRV;
Site 2: NotI; RNA obtained from female animals at 8 wk
old. Tissues were snap-frozen and kept at -80C for two
days before RNA extraction and purification (Tri-reagent
method). cDNA was primed using oligo-dT primer:
5'-pGACTACTTCTAGATCGGCGCGCCGCC(T)25-3' and cloned into
the EcoRV/NotI sites of pExpress-1. Size-selection >1.25kb
resulted in an average insert size of 1.5 kb. This primary
library is normalized (non-normalized primary library is
NIH MGC 252) and was constructed by Express Genomics
(Fredrick, MD). Note: this is a NIH_MGC library."

ORIGIN
Query Match 48.7%; Score 602.6; DB 7; Length 808;
Best Local Similarity 89.6%; Pred. No. 6.6e-131;
Matches 683; Conservative 0; Mismatches 70; Indels 9; Gaps 3;

QY 255 GCGAAGTGTGCAATGCTTATCCAGGAGAAATCAAGAGCTTCGAAGGAGCTGTCCAGGGA 314
DB 14 GCGAAGTGTGCAATGCTTATCCAGGAGAAATCAAGAGCTTCGAAGGAGCTGTCCAGGGA 314
QY 315 GGTATTGACAACTTGAAGCAGAGTTCGGCTTTACAGAGAAATACAGTTCAGGAGCTTAAA 374
DB 74 GGTATTGACAACTTGAAGCAGAGTTCGGCTTTACAGAGAAATACAGTTCAGGAGCTTAAA 133
QY 375 AGCAGTCAGGATGAATGATCTTCATCATCCAGCTTGATGAAGTTCAGGAGTTCAGGATG 434
DB 134 AGCAGTCAGGATGAATGATCTTCATCATCCAGCTTGATGAAGTTCAGGAGTTCAGGATG 193
QY 435 CAGAGTACCAATTCATGCTTCAGCAGCAGCTGAGGAGACAGCCAGGAGTTCAGGATG 494
DB 194 CAGAGTACCAATTCATGCTTCAGCAGCAGCTGAGGAGACAGCCAGGAGTTCAGGATG 253
QY 495 TACAGCAGCAGAGTTCAGGAGCTTCGCCCAAGTACAGCAGGAGTTCAGGATTCGAA 554
DB 254 TACAGCAGCAGAGTTCAGGAGCTTCGCCCAAGTACAGCAGGAGTTCAGGATTCGAA 313
QY 555 CCGTAGAAGCAGTCAGAGGCGCAAGTAAAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAG 614
DB 314 CCGTAGAAGCAGTCAGAGGCGCAAGTAAAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAG 373
QY 615 AATGGTACCTTCCTCCCGCAGAGAGCTTCGGGTCTGGATTTTACAGGAGGCGCAACACA 674
DB 374 AATGGTACCTTCCTCCCGCAGAGAGCTTCGGGTCTGGATTTTACAGGAGGCGCAACACA 433
QY 675 ACCGAATGATCTTCCTTCCTCAGGGAAGTGAATTAAGTTCCTCCCAAGCAGTTCAGAG 734
DB 434 CTTGAGATGATCTTCCTTCCTCAGGGAAGTGAATTAAGTTCCTCCCAAGCAGTTCAGAG 493
QY 735 GAGAGAACTGGCAGAGAGTGTGTTTACGTAATCAACTCAGTTCGGGGTATGAAAGT 794
DB 494 GAGAGAACTGGCA--GAGGTAGTGTATACATAAACCCACTCAGTTCGGGGTATGAAAGT 550
QY 795 GTAGACTTCCTCCAGGCGAGTGAATCTCTCACACACCAATCAATGACACAGCTCC 854
DB 551 GTAGACTTCCTCCAGGCGAGTGAATCTCTCACACACCAATCAATGACACAGCTCC 610
QY 855 AGTCATGACCTCAAGAGGAGAAAGCAGTTCAGTGGGAAAGGTAATCGAATCTGGGGTTC 914
DB 611 AGTCATGACCTCAAGAGGAGAAAGCAGTTCAGTGGGAAAGGTAATCGAATCTGGGGTTC 670
QY 915 CGCAGCTTCAGATGGCTTGGACTCAAGTGAATGATACAGGTTTCAGTTTGAATAT 974

```

```

Db 671 CCCCCATGTCAGAAATGGCTTGGACTCAGTGT-AATGTACAGGGGTCAATTTGTATATTT 729
QY 975 TTTTTCAGCAAAATTTTATACAGTGTCAATTTAATTTGGGAGA 1016
Db 730 TTTCAGC-----AATTTTATACAGTGTCAATTTAATGGGGAAGA 766

RESULT 10
BU942029
LOCUS
DEFINITION
BU942029 710 bp mRNA linear EST 18-OCT-2002
IMAGE:6714512 5', mRNA sequence.
ACCESSION
BU942029
VERSION
BU942029.1 GI:24130848
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 710)
NIH-MGC http://mgi.nci.nih.gov/
AUTHORS
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: NCI
cDNA Library Preparation: Michael Brownstein Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM3026 row: i column: 08
High quality sequence stop: 555.
FEATURES
Location/Qualifiers
1..710
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6714512"
/tissue types="mixed (pool of 40 RNAs)"
/lab host="DH10B (Ti-Phage-resistant)"
/clone_lib="NIH_MGC_127"
/notes="vector: pDNR-LIB; Site 1: SfiI (ggccattatggcc);
Site 2: SfiI (ggccgctcgcc); Double-stranded cDNA was
prepared from a pool of 40 cell line polyA+ RNAs (bladder
- 2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon -
4%, connective tissue - 1.4%, eye - 1%, intestine - 2.6%,
kidney - 2.2%, liver - 5.7%, lung - 10.8%, NK-cell -
5.2%, ovary - 4%, pharynx - 2.5%, prostate - 4.3%,
salivary gland - 1.3%, and skin - 2.3%). 5' and 3'
adaptors were used in cloning as follows:
5'-AAGCAGTGGTATCAACGACAGTGGCCATTCAGCCGCGG-3' and
5'-AATCTAGAGCCGCGGCGGCGCATG-dT(30)NN-3'. Full-length
enriched library was constructed using the Clontech
Creator SMART kit and size-selected to contain the 1-2 kb
size fraction (other fractions present in NIH_MGC_126 and
NIH_MGC_128). Library created in the laboratory of T.
Usdin, M.D., Ph.D. (NIMH, NIH). Note: this is a NIH_MGC
Library."

ORIGIN
Query Match 48.1%; Score 595.6; DB 5; Length 710;
Best Local Similarity 97.3%; Pred. No. 2.9e-129;
Matches 638; Conservative 0; Mismatches 14; Indels 4; Gaps 3;

QY 231 AGCCAAACAGGAGAAAGTTAATGGGAGAGTGTTCGAATGCTTATCCAGGAGATCAAGAG 290
DB 45 AGCCAAACAGGAGAAAGTTAATGGGAGAGTGTTCGAATGCTTATCCAGGAGATCAAGAG 104
QY 291 CTTGGAGGAGCAGTGTCCCGAGGAGCTATTCACAACCTTGACAGCAGTTGGCTTTACAG 350
DB 105 CTTGGAGGAGCAGTGTCCCGAGGAGCTATTCACAACCTTGACAGCAGTTGGCTTTACAG 164

```


MEDLINE 22419904
 PUBMED 12532269
 COMMENT
 Contact: Pomp, D
 Department of Animal Science
 University of Nebraska, Lincoln
 Lincoln, NE 68583-0908, USA
 Tel: 402 472 6416
 Fax: 402 472 6362
 Email: dpomp@unl.edu
 Oligo-dr track not found, Not 1 site shown in beginning of sequence
 is likely internal to the message.
 Seq primer: M13 -29
 POLYA=No.

FEATURES
 source
 1..702
 /organism="Sus scrofa"
 /mol_type="mRNA"
 /strain="University of Nebraska, Lincoln Swine Selection
 Lines"
 /db_xref="taxon:9823"
 /clone="UNL-P-FN-cl-a-02-0-UNL"
 /dev_stage="ADULT"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="UNL-P-FN"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not 1; Site 2: Eco RI; The UNL-P-FN
 library is a normalized library representing porcine
 ovarian follicles, ranging between 2.0 to 10.0 mm in
 diameter, collected during 7 days of the follicular phase
 of the pig estrous cycle. This library was derived from
 the library UNL-P-F2. The tag is a string of 5-6
 nucleotides present between the Not I site and the
 oligo-dr track. The library was constructed as described
 by Ronaldo, Lennon and Soares, Genome Research 6:
 791-806, 1996.
 TAG_SEQ=None found"

ORIGIN
 Query Match 47.7%; Score 590.6; DB 4; Length 702;
 Best Local Similarity 91.6%; Pred. No. 4.3e-128;
 Matches 636; Conservative 0; Mismatches 55; Indels 3; Gaps 1;

QY 249 TTAATGCGAAGTGTGCAATGCTTTATCCAGAGAAATCAAGAGCTTGGAGGCGAGTGTCC 308
 DB |||||
 QY 702 TTAATGCGAAGTGTGCAATGCTTTATCCAGAGAAATCAAGAGCTTGGAGGCGAGTGTCC 643
 DB |||||
 QY 309 CAGGAGCTATTGCACAACTTGAAGCAGAGTGGCTTTACAGAGAAATACAGTCAGGAG 368
 DB |||||
 QY 642 CAGGAGCTATTGCACAACTTGAAGCAGAGTGGCTTTACAGAGAAATATAGTAGGAG 583
 DB |||||
 QY 369 CTTAAAGCAGTCAGGATGAATGACTTCATCATCAGCTTGGATGAAGAGTAGAG 428
 DB |||||
 QY 582 CTTAAAGCAGTCAGGATGAATGACTTCATCATCAGCTTGGATGAAGAGTAGAG 523
 DB |||||
 QY 429 GGTATCAGAGTACCATCTTCTAGTTCGACGACAGCTGAAGAGACAGCCAGAGTGTG 488
 DB |||||
 QY 522 GGTATCAGAGTACCATCTTCTAGTTCGACGACAGCTGAAGAGACAGCCAGAGTGTG 463
 DB |||||
 QY 489 GCTCAGTACAGCAGCAGCAGTCTCAGGCTCTGCCCCCAAGTACCAGCAGGAGTACAGCT 548
 DB |||||
 QY 462 GCTCAGTACCAACAGCAGCAGTCTCAAGCCCCAGGCCCCAGCAGCAGGAGT---CGG 406
 DB |||||
 QY 549 TCTGACCTTAGAAGCAGTCAGAGGCCCAAGTAAGACTGTGAGTGTCTGCAACAGCA 608
 DB |||||
 QY 405 TCTGACCTTAGAAGCAGTCAGAGGCCCAAGTAAGACTGTGAGTGTCTGCAACAGCA 346
 DB |||||
 QY 609 CCAAGTAAGTGTAGTCTCTCCGCCAGAGAGAGCTCTGGGTCTGGATTTTACAGAGGAGGCG 668
 DB |||||
 QY 345 CCAAGTAAGTGTAGTCTCTCCGCCAGAGAGAGCTCTGGGTCTGGATTTTACAGAGGAGGCG 286
 DB |||||
 QY 669 AACACAAACGAGATGACTTCTCTTCTCAGGGAATGGTAATAGTCTCTCCAAACAGC 728
 DB |||||
 QY 285 GACACACCGAAGATGACTTCTCTCTCTCCGGGAGCGGGAATAGGCTCTCCAAACAGC 226
 DB |||||

QY 729 TCAGAGGAGAGAACTGGCAGAGGAGGTAGTGTACGTAAATCAACTCAGTGGGGGTAT 788
 DB |||||
 QY 225 TCAGAGGAGAGAACTGGCAGAGGAGGTAGTGTACGTAAATCAACTCAGTGGGGGTAT 166
 DB |||||
 QY 789 GAAAGTGTAGACTCTCCACCGGCGAGTGAAGAACTCTCTCACACCAATCAATCAACACA 848
 DB |||||
 QY 165 GAAAGCTGTAGACTCTCCACCGGCGAGTGAAGAACTCTCTCACACCGCACTCATGACACA 106
 DB |||||
 QY 849 GACTCCAGTCTATGACCTCAAGAGAGAGAAAGCAGTGTAGTGGAAAGGTAATCGAACTGTG 908
 DB |||||
 QY 105 GACTCCAACTCTGACCTCAAGAGAGAGAAACCGTGTAGTGGAAAGGTAGCCGAACTGCG 46
 DB |||||
 QY 909 GGTTCCTCCCGCAGCTTCAGAAATGGCTTGGACTCAA 942
 DB |||||
 QY 45 GGTTCCTCCCGCAGCTTCAGAAATGGCTTGGACTNCA 12
 DB |||||

RESULT 13
 BU200958
 LOCUS
 DEFINITION
 603949240F1 CSEQCHN03 Gallus gallus mRNA linear EST 25-NOV-2002
 sequence.
 BU200958
 ACCESSION
 BU200958.1 GI:25364439
 VERSION
 BU200958
 KEYWORDS
 EST
 SOURCES
 Gallus gallus (chicken)
 ORGANISM
 Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 1 (bases 1 to 880)
 Boardman,P.E., Sanz-Esquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
 Fong,W.R., Rickie,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
 A Comprehensive Collection of Chicken cDNAs
 Curr. Biol. 12 (22), 1965-1969 (2002)
 22335534
 MEDLINE
 PUBMED
 12445392
 COMMENT
 Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology
 (UMIST)
 PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.

FEATURES
 source
 1..880
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="White Leghorn, Hisex"
 /db_xref="taxon:9031"
 /clone="CHEST906p9"
 /tissue_type="whole embryo"
 /dev_stage="20-21"
 /lab_host="DH10B"
 /clone_lib="CSEQCHN03"
 /note="Organ: whole embryo; Vector: pBluescript II KS(+);
 Site 1: EcoRI; Site 2: NotI; This normalized library was
 constructed from 1 million independent clones cDNA
 synthesis was initiated using an oligo(dT) primer, using
 methylated C in the first strand synthesis reaction.
 Following this first strand reaction, double-stranded cDNA
 was bluntended, ligated to NotI adapters, digested with
 EcoRI, size-selected, and cloned into the NotI and EcoRI
 compatible sites of a custom modified MCS of the
 pBluescript (KS+) vector. The library was normalized in 2
 rounds using conditions adapted from Soares et al., PNAS
 (1994) 91: 9228-9232 and Bernaldo et al., Genome Research 6
 (1996): 791, except that a significantly longer
 reannealing hybridization was used."

ORIGIN

Query Match 46.2%; Score 571.8; DB 5; Length 880;
Best Local Similarity 84.2%; Pred. No. 1.2e-123; Indels 1; Gaps 1;
Matches 656; Conservative 0; Mismatches 122;

273 ATCCAGAGAAATCAAGAGCTTGAAGCAGCTCTCCAGGAGCGTATTCAGCAACTTGAA 332
Db 1 ATCCAGAGAAATCAAGAGCTTGAAGCAGCTCTCCAGGAGCGTATTCAGCAACTTGAG 60

333 GCAGAGTTGCTTTACAGAGAAATACAGTGAGAGCTTAAAGCAGCTCAGGATGAAC 392
Db 61 GCAGAGTTGCTTTACAGAGAAATACAGTGAGAGCTTAAAGCAGCTCAGGATGAATG 120

393 AATGACTTTCATCCAGCTTGATGAAGAAAGTAGAGGTTATGCAGAGTACCATTTAGTT 452
Db 121 AATGACTTTCATCCAGCTTGATGAAGAGGTAGAGGTTATGCAGAGTACCATTTAGTG 180

453 CTGCAGAGAGCTGAAGAGAGACAGCCAGAGTTGGCTCAGTACCAGCAGCAGCTCT 512
Db 181 CTTCAGCAGAGTTGAAGAGAGCTCCGCGAGCTTGGCGAGTACCAGCAGCAGCTCC 240

513 CAGGCTCTGCCCAAGTACAGCAGAGCTACAGCTTCTGAACCTGTAGAACAGTCAGAG 572
Db 241 CAGGCTCTCCACCGAGTACAGCAGAGCTCCATCTTCGAGCTACAGACCGAGGAG 300

573 GCCAAGATGAAGACTCGCTCTGACAAACGAGCAAGTAAAGTAGTCTCTCTCCCG 632
Db 301 GCTGTGGTAAAGACTCAGCGCTGTGGCTAAACGAGCAAGCAATGGTAGCTCTCTCCAT 360

633 CAGAGGAGCTCTGGGTCTGGATTTACAGGAGGAGCAACACCGAAGATGACTTTCCT 692
Db 361 CAGGAGAGCTCTGGGCTGGATTTATAGGAGGAGTACGACAGAGATGACTTCCCG 420

693 TCTTCTCCAGGAAATGTAATAGTCTCTCAACAGCTCAGAGGAGAGAACTGCGAGGA 752
Db 421 GCTTCTCCAGGAAATGTAATAGCTGTCCAAACCACTCTGAAGATAGAACTGGTAGGA 480

753 GGTAGTGGTTACGTAATCAACTCAGTTCGGGGTATCAAAAGTGTAGACTCTCCACGGC 812
Db 481 GGTGTAGTACATATAAACCACTCAGTACTGGGTATGAAAGTGTAGACTCTCCACTGGC 540

813 AGTGAAACTCTCTCACACCAATCAATGACACAGCTCCAGTCAATGACCTTCAAGAG 872
Db 541 AGTGAAACTCTCTCACTCCACTCAATGACACAGCTCCCAATCATGATCTCTCAAGAG 600

873 GAGAAAGCAGTGTGGGAAAGTAACTGAACTGTGGTTCGCCCGCAGCTTCAGAAATGGC 932
Db 601 GAGAAACAGTAAGCATGAAGAGTAAACAGAACTCGGGTTCTCGGCATGTCCAGAAATGGT 660

933 TTGGACTCAAGTGAATGTACAGGTTTCAGTTTGTAAATATTTTTCAGCAATTTTA 992
Db 651 TTGGACTCAATGATACATGTCAGGCTTCAGTTTGTAAATATTTTTCGAGCATTTTAT 720

993 TACAGTGTCAATTAATTTGGAGA-GGATACGTCCAGAAATTAATGCATACCTTTGT 1050
Db 721 ACAGTGTCAATTAATAGGAGAGGAGTACTGTTCAGAAAGCGCTTATTAACATAGT 779

RESULT 14
AJ657839 660 bp mRNA linear EST 28-JUN-2004
LOCUS AJ657839 KN277 Sus scrofa cDNA clone C0005211_H09, mRNA sequence.
DEFINITION
ACCESSION AJ657839
VERSION AJ657839.1 GI:49341908
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 660)
Anderson, S. I., Finlayson, H. A. and Archibald, A. L.
Development of cDNA and EST resources for studying reproduction and
embryo development in pigs and cattle
Unpublished (2004)

COMMENT

Contact: Anderson SI
Genomics and Bioinformatics
Roslin Institute
Roslin, Midlothian, EH25 9PS, UNITED KINGDOM
Single pass sequencing. Bases called and trimmed with phred
v0.020425.c. Vector identified by cross match with the -minscore 20
and -minmatch 12 options. Vector:pBlueScriptII (SK+) R. Site1: EcoRI
R. Site2: NotI 5' Seq Primer M13F Normalised library constructed
from pooled early embryos, from 8- cell stage to blastocysts.
Clones available from UK Centre for Functional Genomics in Farm
Animals, Roslin Institute, Roslin, Midlothian, UK, EH25 9PS,
www.arkgenomics.org.

Location/Qualifiers
1..660
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/clone="C0005211_H09"
/tissue_type="embryo"
/clone_lib="KN277"

FEATURES
source

ORIGIN

Query Match 45.8%; Score 566.4; DB 1; Length 660;
Best Local Similarity 92.5%; Pred. No. 2.1e-122;
Matches 607; Conservative 0; Mismatches 46; Indels 3; Gaps 1;

200 TCAGTCTGAGTCTGCTGAAAGTTTACGCTGATAGCCAAACAGGAAAGTTAATGGCGAA 259
Db 4 TGAACGTGAGTCTGCTGAAAGTTTACGCTGATAGCCAAACAGGAAAGTTAATGGCGAA 63

260 GTGTGCAATGCTTATCCAGGAGAAACAAGAGCTTGAAGAGGAGCTGTCCAGGGACGAT 319
Db 64 GTGTGCAATGCTTATCCAGGAGAAACAAGAGCTTGAAGAGGAGCTGTCCAGGGACGAT 123

320 TGCACAACTTGAAGCAGAGTTGGCTTTACAGAGAAATACAGTGAAGAGCTTAAAGCAG 379
Db 124 TGCACAACTTGAAGCAGAGTTGGCTTTACAGAGAAATATAGTGAAGAGCTTAAAGCAG 183

380 TCAGATGAATCAATGACTTCATCATCCAGCTTGAAGAGGAGTAGAGGTTATCGAGAG 439
Db 184 TCAGATGAATCAATGACTTCATCATCCAGCTTGAAGAGGAGTAGAGGTTATCGAGAG 243

440 TACCAATCTAGTCTTGACAGCAGCTGAGAGAGACACCCAGCAGTGGCTCAGTACCA 499
Db 244 TACCAATCTAGTCTTCAACGGCAGCTGAAGAGAGACCCGCCAGCAGCTGGCTCAGTACCA 303

500 GCAGCAGCAGTCTCAGGCTCTGCCCAAGTAGTACCAGCAGGACTACAGCTTCTGAACCTGT 559
Db 304 ACAGCAGCAGTCTCAAGCCCCAGGCCGACACAGCAGGACT---CCGCTCTGAGCCTGG 360

560 AGAACAGTCAGAGGCCAACAGTAAAGACTGCAAGTCTGACAAACGGACCAAGTAATGG 619
Db 361 AGAACAGTCAGAGGCCAACAGGGAAGACTGCAAGTCTGCGCAACGGACCAAGTAATGG 420

620 TAGCTCTCTCCGCGCAGAGAGCGCTGGGTCTGGATTTACAGGAGGAGGCAACACACCA 679
Db 421 CAGCTCTCTCCGCGCAGAGAGCGCTGGGTCTGGGTTCAGGAGGAGGCGACACAGCAGCA 480

680 AGATGACTTTTCCTTCTTCTCCAGGGAATGTAATAGTCTTCCAAACAGCTCAGAGAGAG 739
Db 481 AGATGACTTTTCCTTCTTCTCCGCGGAGCGGGAATAGGCTTCCAAACAGCTCAGAGAGAG 540

740 AACTGCGACAGAGGAGTAGTGGTTAGTAAATCAACTCAGTGGGGGTATGAAGTGTAGA 799
Db 541 AACTGCGACAGAGGAGTAGTGGTTAGTAAATCAACTCAGTGGGGGTATGAAGTGTAGA 600

800 CTCTCCACCGGGCAGTGAAACCTCTCTCACACCAACCAATCAAATGACACAGACTCCA 855
Db 601 CTCTCCACCGGGCAGTGAAACCTCTCTCACACCGGCACTCACATGACACAGACTCCA 656

RESULT 15	CB182874	961 bp	mRNA	linear	EST 16-MAY-2003
LOCUS	AGENCOURT 11381782 NIH MGC 164 Mus				
DEFINITION	IMAGE:30243672 5', mRNA sequence.				
ACCESSION	CB182874				
VERSION	CB182874.2	GI:291133999			
KEYWORDS	Est.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	1 (bases 1 to 961)				
TITLE	NIH-MGC http://mgi.nci.nih.gov/.				
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)				
COMMENT	Unpublished (1999)				
	On Jan 30, 2003 this sequence version replaced gi:28181082.				
	Contact: Robert Strausberg, Ph.D.				
	Email: cgabs-x@mail.nih.gov				
	Tissue Procurement: Dr. David Rowe and Dr. Mina				
	cDNA Library Preparation: Invitrogen Corp				
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)				
	DNA Sequencing by: Agencourt Bioscience Corporation				
	Clone distribution: MGC clone distribution information can be				
	found through the I.M.A.G.E. Consortium/LNL at:				
	http://image.llnl.gov				
	Plate: NDAM0315 row: f column: 01				
	High quality sequence start: 50				
	High quality sequence stop: 696.				
FEATURES	Location/Qualifiers				
source	1..961				
	/organism="Mus musculus"				
	/mol_type="mRNA"				
	/db_xref="taxon:10090"				
	/clone="IMAGE:30243672"				
	/lab_host="DH10B (phage-resistant)"				
	/clone_lib="NIH MGC 164"				
	/notes="Vector: pCW-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Non-normalized full-length enriched library from Pooled mouse embryonic limb, maxilla and mandible, day 10.5 and 11.5 (size selected for the 0.5-1 kb fragments) Cloned directionally, priming method: Oligo-dr. cDNA enrichment: >1x bp. Average insert size 1.8k bp. Priming sequence: 5'GACTGCTCTAGATCGGAGCGGCCCTTT 3'. Tissue contributed by, David Rowe. Library constructed by ResGen, Invitrogen Corp."				
ORIGIN					
Query Match	44.7%;	Score 553.4;	DB 6;	Length 961;	
Best Local Similarity	85.5%;	Pred. No. 2.6e-119;			
Matches	678;	Conservative 0;	Mismatches 76;	Indels 39;	Gaps 4;
QY	447	CTAGTTCTGCAGCAGCGTCTGAAGGAGACACGCGCAGCAGTTGGCTCAGTACCAGCAGCAG 506			
Db	3	CTGGTACCGTCCGGAATCTTCGGGATGAGCGACAGCAGTTGGCACATTTCCCGCAGCAG 62			
QY	507	CAGTCTCAGGCTCTGCCCCAGTACGAGACCTACAGCTTCTGAACTGTAGAACAG 566			
Db	63	CAGTCTCAAGTTTCAGCTCCAAGTACAGCAGGACTACATCTTCTGAACTGTAGATCAG 122			
QY	567	TCAGGCGCCACAAGTAAAGACTGCAGTCGTTGTACAAACGGACCAAGTAATGATGCTCC 626			
Db	123	GCAGAGGTCACAAGCAAGACTGCAGTCGTTGGCAATGGACCAAGTAATGCGACTCC 182			
QY	627	TCCCGCCAGAGACGTCCTGGTCTGGATTTCACGGGAGGGAACACACCGAAGTAC 686			
Db	183	TCCCGCCAGAGACGTCCTGGTCTGGATTTCACGGGAGGAGACACCTGAGGATGAC 242			
QY	687	TTTCTCTTTCTCCAGGGAATGGTAATAAGTCCTCAACAGCTCAGAGGAGAACTGGC 746			
Db	243	TTTCTCTTTCTCCAGGGAATGGTAATAAGGCTCCACAGCTCAGAGGAGAACTGGC 302			

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2004, 12:15:48 / Search time 191 Seconds
(without alignments)
960.966 Million cell updates/sec

Title: US-10-030-389-2
Perfect score: 1608
Sequence: 1 MRLATKGQDIQECTTQIQL.....GSRHVQGLDSSVNVQGSVL 319

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt 02.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1593	99.1	396	1	WTAP_HUMAN
2	1593	99.1	396	1	Aah69192 homo sapi
3	1531	95.2	396	1	WTAP_MOUSE
4	1296.5	80.6	393	2	Q6P4K5
5	1296.5	80.6	393	2	Aah63362 xenopus t
6	1146.5	71.3	423	2	Q7SXL7
7	528.5	32.9	412	2	Q7KA73
8	528.5	32.9	539	2	Q9Y091
9	525.5	32.7	412	2	Q8ML74
10	525.5	32.7	536	2	Q9V6T8
11	518	32.2	272	2	Q7Q0B0
12	357	22.2	74	2	Q9BZS4
13	282	17.5	330	2	Q9ZS28
14	218	13.6	267	2	Q6PNZ9
15	218	13.6	267	2	Q7SXL7
16	190.5	11.8	554	2	Q7RP86
17	172.5	10.7	617	2	Q8T5B2
18	168	10.4	652	2	Q7KXV9
19	168	10.4	652	2	Aa338587
20	166	10.3	524	2	Q9K2U9
21	165.5	10.3	576	2	Q9KJ33
22	163	10.1	114	2	Q7Q0B1
23	160	10.0	943	2	Q7QD47
24	158.5	9.9	692	2	Q6TPG6
25	158.5	9.9	692	2	Aa094949
26	158.5	9.9	725	2	Q8MPY1
27	158.5	9.9	725	2	CA044152
28	158.5	9.9	1509	2	Q7JN47
29	158.5	9.9	1509	2	CA031488
30	158.5	9.9	1888	2	Q7JN45
31	158.5	9.9	1888	2	CA031490

32	158.5	9.9	1927	2	Q7JN46
33	158.5	9.9	1927	2	CA031489
34	158.5	9.9	2048	2	Q7JN42
35	158.5	9.9	2048	2	CA031493
36	158.5	9.9	2101	2	Q7JN43
37	158.5	9.9	2101	2	CA031492
38	158.5	9.9	2134	2	Q7JN44
39	158.5	9.9	2134	2	CA031491
40	157	9.8	995	2	Q9NJ19
41	157	9.8	1229	2	Q9NJ22
42	157	9.8	1243	2	Q9NJ21
43	157	9.8	1253	2	Q9NJ20
44	157	9.8	1951	2	Q17042
45	156.5	9.7	516	1	P54_ENTFC

ALIGNMENTS

RESULT 1
WTAP_HUMAN STANDARD; PRT; 396 AA.
ID WTAP_HUMAN Q9BYJ7; Q9H4E2;
AC Q15007; Q9GT28; Q9BYJ7; Q9H4E2;
DT 15-JUL-1998 (Rel. 36, Created)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Wilms' tumor 1-associated protein (WT1-associated protein) (Putative
DE pre-mRNA splicing regulator female-lethal (2D) homolog).
DE Name=WTAP; Synonyms=KIAA0105;
GN Homo sapiens (Human)
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Ortega A., Penalva L., Valcarcel J.;
RT "cDNA of human homolog of the drosophila putative pre-mRNA splicing
RT regulator female-lethal (2D) [FL(2D)]";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Bone marrow;
RX MEDLINE=95308325; PubMed=7788527;
RA Negase T., Miyajima N., Tanaka A., Sazuka T., Seki N., Sato S.,
RA Tabata S., Ishikawa K.-I., Kawabayashi Y., Kotani H., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. III.
RT The coding sequences of 40 new genes (KIAA0081-KIAA0120) deduced by
RT analysis of cDNA clones from human cell line KG-1.";
RL DNA Res. 2:37-43(1995).
[3]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Lung;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.F., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Prange C.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Patrino L.,
RA Raha S.A., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting J., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."; ;
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

Q7Jm46 caenorhabdi
Caf31489 caenorhab
Q7Jm42 caenorhabdi
Caf31493 caenorhab
Q7Jm43 caenorhabdi
Caf31492 caenorhab
Q7Jm44 caenorhabdi
Caf31491 caenorhab
Q9nj19 asquiptecten
Q9nj22 asquiptecten
Q9nj21 asquiptecten
Q9nj20 asquiptecten
Q17042 asquiptecten
P13692 enterococcu

```

RN [4]
RP SEQUENCE OF 4-391 FROM N.A. (ISOFORM 1), INTERACTION WITH WTI,
RP SUBCELLULAR LOCATION, AND TISSUE SPECIFICITY.
RX TISSUE=Fetal Kidney;
RX MEDLINE=20458888; PubMed=11001926;
RA Little N.A., Hastie N.D., Davies R.C.;
RT Identification of WTAP, a novel Wilms' tumour 1-associating
RL protein."
RL Hum. Mol. Genet. 9:2231-2239(2000).
RN [5]
RP SEQUENCE OF 112-396 FROM N.A. (ISOFORM 1).
RC TISSUE=Myosin;
RA Wambutt R., Heubner D., Mewes H.-W., Weil B., Wiemann S.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SUBUNIT: Interacts with WTI.
CC -!- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q15007-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q15007-2; Sequence=VSP_010278, VSP_010279;
CC -!- TISSUE SPECIFICITY: Ubiquitous.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AF374416; AAK34764.1; -
DR EMBL; D14661; BAA03495.1; -
DR EMBL; BC000383; AAK00383.1; -
DR EMBL; BC004432; AAK04432.1; -
DR EMBL; AJ276706; CAC10188.1; -
DR EMBL; AL583911; CAC29495.1; -
DR EMBL; GenBank; HGNC:16846; WTAP.
DR MIM; 605442; -
KW Alternative splicing; Nuclear protein.
FT VARSPLIC 151
FT FTID=VSP_010278.
FT VARSPLIC 152
FT FTID=VSP_010279.
FT FTID=VSP_010279.
SQ SEQUENCE 396 AA; 44243 MW; B4F81B3000P77A37 CRC64;

Query Match 99.1%; Score 1593; DB 1; Length 396;
Best Local Similarity 99.1%; Pred. No. 1.le-72;
Matches 316; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MRLATKQDQIECTTQIQLKQVQPSVAQLRSTWDPAINLFLKMKGELEOTKDKLEQ 60
Db 78 MRLATKEQEQEECTTQIQLKQVQPSVAQLRSTWDPAINLFLKMKGELEOTKDKLEQ 137
Qy 61 AQNELSAWKTTPDSQTKLMAKCRMLIQENQELGRLSQGRIQAQLAEALQKYSSEL 120
Db 138 AQNELSAWKTTPDSQTKLMAKCRMLIQENQELGRLSQGRIQAQLAEALQKYSSEL 197
Qy 121 KSSQDELNDPFIQLDDEVEGMQSTILVLOQLKETRQQLAQYQQQSQASAPSTRTTAS 180
Db 198 KSSQDELNDPFIQLDDEVEGMQSTILVLOQLKETRQQLAQYQQQSQASAPSTRTTAS 257
Qy 181 EPVEQSEATSKDCSRLTNGPSNGSSSRQRTSGSGFHREGNTTDDPFPSPGNGKSNSS 240
Db 258 EPVEQSEATSKDCSRLTNGPSNGSSSRQRTSGSGFHREGNTTDDPFPSPGNGKSNSS 317
Qy 241 EERTGRGGSGYVNLQAGYESVDSPTGSENSTHQSNDTSSHDPOEEKAVSGKGNRTVG 300
Db 318 EERTGRGGSGYVNLQAGYESVDSPTGSENSTHQSNDTSSHDPOEEKAVSGKGNRTVG 377
Qy 301 SRHVQNLGDSVNVQGSVL 319

```

```

Db 378 SRHVQNLGDSVNVQGSVL 396
RESULT 2
AAH69192
ID AAH69192 PRELIMINARY; PRT; 396 AA.
AC AAH69192;
DT 10-MAY-2004 (TREMELREL. 27; Created)
DT 10-MAY-2004 (TREMELREL. 27; Last sequence update)
DT 10-MAY-2004 (TREMELREL. 27; Last annotation update)
DE Wilms' tumour 1-associating protein, isoform 1.
GN WTAP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson K.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.,
RA Krzywinski M.I., Skalska J., Schmutz J., Myers R.M., Butlerfield Y.S.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Strausberg R.;
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC069192; AAH69192.1; -
SQ SEQUENCE 396 AA; 44243 MW; B4F81B3000P77A37 CRC64;

```

```

Query Match 99.1%; Score 1593; DB 2; Length 396;
Best Local Similarity 99.1%; Pred. No. 1.le-72;
Matches 316; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MRLATKQDQIECTTQIQLKQVQPSVAQLRSTWDPAINLFLKMKGELEOTKDKLEQ 60
Db 78 MRLATKEQEQEECTTQIQLKQVQPSVAQLRSTWDPAINLFLKMKGELEOTKDKLEQ 137
Qy 61 AQNELSAWKTTPDSQTKLMAKCRMLIQENQELGRLSQGRIQAQLAEALQKYSSEL 120
Db 138 AQNELSAWKTTPDSQTKLMAKCRMLIQENQELGRLSQGRIQAQLAEALQKYSSEL 197
Qy 121 KSSQDELNDPFIQLDDEVEGMQSTILVLOQLKETRQQLAQYQQQSQASAPSTRTTAS 180
Db 198 KSSQDELNDPFIQLDDEVEGMQSTILVLOQLKETRQQLAQYQQQSQASAPSTRTTAS 257
Qy 181 EPVEQSEATSKDCSRLTNGPSNGSSSRQRTSGSGFHREGNTTDDPFPSPGNGKSNSS 240
Db 258 EPVEQSEATSKDCSRLTNGPSNGSSSRQRTSGSGFHREGNTTDDPFPSPGNGKSNSS 317
Qy 241 EERTGRGGSGYVNLQAGYESVDSPTGSENSTHQSNDTSSHDPOEEKAVSGKGNRTVG 300
Db 318 EERTGRGGSGYVNLQAGYESVDSPTGSENSTHQSNDTSSHDPOEEKAVSGKGNRTVG 377
Qy 301 SRHVQNLGDSVNVQGSVL 319

```



```

DB 378 SRHVQGLDSSVNVQGSVL 396
|||||
RESULT 3
WTAP_MOUSE STANDARD; PRT; 396 AA.
AC Q9ER69;
DT 05-JUL-2004 (Rel. 44, Created)
DT 05-JUL-2004 (Rel. 44, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Wtaps' tumor 1-associating protein (Wtl-associating protein) (Putative
DE pre-mRNA splicing regulator female-lethal(2D) homolog).
GN Names=Wtaps;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE OF 18-396 FROM N.A. (ISOFORM 1).
RX MEDLINE=20458888; PubMed=11001926;
RA Little N.A., Hastie N.D., Davies R.C.;
RT "Identification of WTAP, a novel Wilms' tumour 1-associating
RT protein."
RL Hum. Mol. Genet. 9:2231-2239(2000).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC STRAIN=C57BL/6J;
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Baidarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt T.A., Brusci V., Chochia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawai H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Nunata K., Okido T., Pavan W.J., Pettea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., King B.Z., Ringwald M.,
RA Sadelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA Sultana R., Takemura Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo L., Wagner L., Wahlstedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Waki J., Alzawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC STRAIN=Czech II; TISSUE=Breast tumor;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieff P.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Spatlenko M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustun T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullah S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

```

```

RA Fahey J., Helton E., Kerteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Sutterfield Y.S.N., Krzywinski M.I., Skalska U., Smalish D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- SUBUNIT: Interacts with Wtl (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear; nucleolar (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q9ER69-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9ER69-2; Sequence=VSP_010280, VSP_010281;
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ276707; CAC16790.1; -
DR EMBL; AK076111; BAC36191.1; -
DR EMBL; BC046416; AAH46416.1; -
DR MGD; MGI:1926395; Wtaps.
KW Alternative splicing; Nuclear protein.
FT VARSPPLIC 151 151 S -> R (in isoform 2).
FT VARSPPLIC 152 396 Missing (in isoform 2).
FT SEQUENCE 396 AA; 44215 MW; 63900F41F9D01777 CRC64;
SQ
Query Match 95.2%; Score 1531; DB 1; Length 396;
Best Local Similarity 94.7%; Pred. No. 1.5e-69;
Matches 302; Conservative 8; Mismatches 9; Indels 0; Gaps 0;
Qy 1 MLATKGQDIQECTTQYLYKQVQPSVAQLRSTMDPAINLFLKMKGELEQTKKLEQ 60
Db 78 MLATKKEQEMQECTTQYLYKQVQPSVAQLRSTMDPAINLFLKMKGELEQTKKLEQ 137
Qy 61 AQNELSANKFPDSTGKLMKCRMLIQNELGRQLSQGRIQLAEALQKYSSEEL 120
Db 138 AQNELSANKFPDSTGKLMKCRMLIQNELGRQLSQGRIQLAEALQKYSSEEL 197
Qy 131 KSSQDELNDFFIQLDDEVEGQSTILVQLQOLKETROQLAQYQQQSQASAPSTRTAS 180
Db 198 KSSQDELNDFFIQLDDEVEGQSTILVQLQOLKETROQLAQYQQQSQASAPSTRTSS 257
Qy 181 EPVQSEATSKDCSRLTNGPSNGSSSRQRTSGSGFHREGNTTDDFPSSPGNGKSNSS 240
Db 258 EPVQAEVTSKDCSRLANGPSNGSSSRQRTSGSGFHREGSTPEDDPSSSGNGKSNSS 317
Qy 241 ERTGCGSGYVQLSAGYSDVPTGSENLSHQSDNDTSSHPQEKAVSGKGRVTG 300
Db 318 EERTGCGSSYINPLSAGYSDVPTGSENLSHQSDNDTSSHPQEKAVSGKGRVTG 377
Qy 301 SRHVQGLDSSVNVQGSVL 319
Db 378 SRHVQGLDSSVNVQGSVL 396
RESULT 4
Q6P4K5 PRELIMINARY; PRT; 393 AA.
ID Q6P4K5;
AC Q6P4K5;
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)

```

```

DE Hypothetical protein MGC75954.
GN Name=MGC75954;
OS Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8364;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whitling M., Touchman J.W., Green E.D., Dickinson M.C.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalhus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Klein S., Strausberg R.;
RA Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC063362; AAH63362.1; -.
DR InterPro; IPR009053; Prefoldin.
KW Hypothetical protein.
SQ
SEQUENCE 393 AA; 44312 MW; 82983BAB02EA99AE CRC64;

Query Match 80.6%; Score 1296.5; DB 2; Length 393;
Best Local Similarity 81.2%; Pred. No. 9.2e-58;
Matches 260; Conservative 23; Mismatches 32; Indels 5; Gaps 2;

QY 1 MRLATKGODIOECTTQIOLYKQVQPSVAQLRSTWMDPAINLFLKXKGLSEOTKDLQ 60
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
78 MRLATKEQMEQECTTQIHLKQVQPSVAQLRATWMDPAINLFFIKKAELEQTKDLQ 137
QY 61 AQNELSAWKFTPDSTQTKLMAKCRMLIQENQELGRLSQGRIAEALALQKYSSEEL 120
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
138 AQNELSAWKFTPDSTQTKLMAKCRMLIQENQELGRLSQGRIAEALALQKYSSEEL 197
QY 121 KSSODELNDFFIQLDEVEGQSTILVLOOLKETROQLAQYQQQSQASAPSTRTAS 180
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
198 KSSODELNDFFIQLDEVEGQSTILVLOOLKDSRQLSQFQQQ----IQ*SGNRTPS 253
QY 181 EPVQSEATSKDCSRLNPGNSGSSRQRTSGS-GFHRENTTDDPPSPGNGKSSNS 239
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
254 ESKDEGTSKDCGRILNPGNSGSSHQRTSHSVGLYREGSSTEEDFASPINEGKLPNH 313
QY 240 SEERTGGGGYNNQLSAGYESVDSPGSENSLTHOSNDTDSHDPQEEKAVSGKGNRTV 299
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
314 SEERTSGGSSYNNQLSTGYESVDSPGSENSLTHOSNDTDSHDPQEEKAVSGKGNRTV 373
QY 300 GSRHVQNGLDSSVNVQGSVL 319
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:
374 SSRHLQNGLDSSVNVQGSVL 393

RESULT 5
AAH63362
ID AAH63362 PRELIMINARY; PRT; 393 AA.

```

```

AAH63362;
25-MAR-2004 (T-EMBLrel. 27, Created)
25-MAR-2004 (T-EMBLrel. 27, Last sequence update)
25-MAR-2004 (T-EMBLrel. 27, Last annotation update)
DE Hypothetical protein MGC75954.
GN MGC75954.
OS Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus; Silurana.
OX NCBI_TaxID=8364;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whitling M., Touchman J.W., Green E.D., Dickinson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalhus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Klein S., Strausberg R.;
RA Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC063362; AAH63362.1; -.
KW Hypothetical protein.
SQ
SEQUENCE 393 AA; 44312 MW; 82983BAB02EA99AE CRC64;

Query Match 80.6%; Score 1296.5; DB 2; Length 393;
Best Local Similarity 81.2%; Pred. No. 9.2e-58;
Matches 260; Conservative 23; Mismatches 32; Indels 5; Gaps 2;

QY 1 MRLATKGODIOECTTQIOLYKQVQPSVAQLRSTWMDPAINLFLKXKGLSEOTKDLQ 60
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
78 MRLATKEQMEQECTTQIHLKQVQPSVAQLRATWMDPAINLFFIKKAELEQTKDLQ 137
QY 61 AQNELSAWKFTPDSTQTKLMAKCRMLIQENQELGRLSQGRIAEALALQKYSSEEL 120
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
138 AQNELSAWKFTPDSTQTKLMAKCRMLIQENQELGRLSQGRIAEALALQKYSSEEL 197
QY 121 KSSODELNDFFIQLDEVEGQSTILVLOOLKETROQLAQYQQQSQASAPSTRTAS 180
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
198 KSSODELNDFFIQLDEVEGQSTILVLOOLKDSRQLSQFQQQ----IQ*SGNRTPS 253
QY 181 EPVQSEATSKDCSRLNPGNSGSSRQRTSGS-GFHRENTTDDPPSPGNGKSSNS 239
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
254 ESKDEGTSKDCGRILNPGNSGSSHQRTSHSVGLYREGSSTEEDFASPINEGKLPNH 313
QY 240 SEERTGGGGYNNQLSAGYESVDSPGSENSLTHOSNDTDSHDPQEEKAVSGKGNRTV 299
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
314 SEERTSGGSSYNNQLSTGYESVDSPGSENSLTHOSNDTDSHDPQEEKAVSGKGNRTV 373
QY 300 GSRHVQNGLDSSVNVQGSVL 319
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:
374 SSRHLQNGLDSSVNVQGSVL 393

```

```
RESULT 6
Q7SXL7 PRELIMINARY; PRT; 423 AA.
ID Q7SXL7
AC Q7SXL7; (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE Wlms; tumour 1-associating protein.
GN Name=wtap;
OS Brachydanio rerio (zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OC NCBI_TaxID=7955;
CX [1]
RN R1
RP SEQUENCE FROM N.A.
TX TISSUE=whole body; PubMed=12477932;
RX MEDLINE=22389257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heien F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalusz D.B., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
TX TISSUE=whole body;
RA Strausberg R.;
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC055544; AAH55544.1; -.
SQ SEQUENCE 423 AA; 46151 MW; A2600E687DAA229C CRC64;

Query Match 71.3%; Score 1146.5; DB 2; Length 423;
Best Local Similarity 70.9%; Pred. No. 3.5e-50; Mismatches 25; Indels 29; Gaps 8;
Matches 246; Conservative 25;

QY 1 MRLATKGDIQECTTQIQLKQVQPSVAQLRSTWVDPAINLFFLKMKGELEQTKKLEQ 60
Db 1 MRLATKEQMEQECTTQIQLKQVQPSVAQLRSTWVDPAINLFFLKMKGELEQTKKLEQ 137
QY 61 AQNELSAWKFTPDSTQTKLMKACRMILQENQELGRQLSQRIQLAEALQKYSFEL 120
Db 138 AQNELSAWKFTPDSTQTKLMKACRMILQENQELGRQLSQRIQLAEALQKYSFEL 197
QY 121 KSSQDELNDFFIQLDEEVEGQSTILVLQOQLKETROQLAQYQQQSQASAPTSRT--- 177
Db 198 KSSQDELNDFFIQLDEEVEGQSTILVLQOQLKETROQLAQYQQQSQASAPTSRTSPS 257
QY 178 TASEPVEVESEATS-----KDCSLTNGPNSGSSRRQTSRGPHRGNTTDDFFPSP 230
Db 258 TASEPSTQEPANASSNVKDCGRVNSGPNSSRGSGSSLYREASSADEYFPSP 317
QY 231 G-----NG-NKSSNSSSEERTG-RGSGGVNQLSAGYSDSPGTSNSLTHQSNDDTSS 282
Db 318 SVSSPTHGDISKLSNHSNDAVQRCGEGYVITQLSAGYSDSPGTSNTHQSNDDTSSN 377
QY 283 HDPQEKAVSGKGNRTVGRSH-VQGLDSSVNV-----QGSVL 319
Db 378 ADSHEAAAAP-KGSRTAGTRHSTQNGLDSSAAVAATNTSNASAGSVL 423
```

```
RESULT 7
Q7KA73 PRELIMINARY; PRT; 412 AA.
ID Q7KA73
AC Q7KA73; (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE FL(2)D-s protein.
GN Name=fl(2)d;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_TaxID=7227;
RN R1
RP SEQUENCE FROM N.A.
TX TISSUE=whole body; PubMed=10790389;
RX MEDLINE=20253098; PubMed=10790389;
RA Penalba L.O.F., Ruiz M.F., Ortega A., Granadino B., Vicente L.,
RA Segarra C., Valcarcel J., Sanchez L.;
RT "The Drosophila fl(2)d gene, required for female-specific splicing of
RT Sxl and tra pre-mRNAs, encodes a novel nuclear protein with a HQ-rich
RT domain."
RL Genetics 155:129-139(2000).
DR EMBL; AJ243607; CAB46726.1; -.
SQ SEQUENCE 412 AA; 45055 MW; 17979281B158D581 CRC64;

Query Match 32.9%; Score 528.5; DB 2; Length 412;
Best Local Similarity 33.9%; Pred. No. 4.3e-19; Mismatches 62; Indels 73; Gaps 8;
Matches 127; Conservative 62;

QY 2 RLATKGDIQECTTQIQLKQVQPSVAQLRSTWVDPAINLFFLKMKGELEQTKKLEQ 61
Db 18 RLANKKEQEPQDVVSQIAEYKAAQAPALALRALDPAVNLFFERLKKELKATKALLET 77
QY 62 QNELSAWKFTPDSTQTKLMKACRMILQENQELGRQLSQRIQLAEALQKYSFEL 121
Db 78 QNELSAWKFTPDSTQTKLMKACRMILQENQELGRQLSQRIQLAEALQKYSFEL 137
QY 122 SSQDELNDFFIQLDEEVEGQSTILVLQOQLKETROQLAQYQQQSQ-----ASA 171
Db 138 KQSELDLDFLQELDEDVEGQSTILFLQOQLKETTRDRIQLEKNAQLKQAIKDEVADA 197
QY 172 PSTRRTTASEPVEQSEATSKDCSLTNGPNSGSSRRQTSRGPHRGNTTDDFFPSS-- 229
Db 198 AATNGGT-NTTINKLETIHEDAC-NANNPTNDC-----YNGNTNNEQIAAAPQ 244
QY 230 --PGNGKSSNSSEERTGSGGVNQLSAGYSDSPGTSNSL----- 272
Db 245 IPPSDGSGNNNGVAARLAKRNYQEEALPTVVVPTPTPGVNNVQEAAPPREVTAPTL 304
QY 273 -----THQSNDDTSSHDPOB-----EKAVSGKGNRTVGRSHV----- 304
Db 305 PPKSKLGRITTRNSQLEDEHQVTPVAVPMVDNAVAGMASEEAAAAAANNVNTG 364
QY 305 ---QGLDSSVNVQ 316
Db 365 IIPETGVQGVPEV 379
```

```
RESULT 8
Q9Y091 PRELIMINARY; PRT; 539 AA.
ID Q9Y091
AC Q9Y091; Q9U974;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE FL(2)D protein.
GN Name=fl(2)d;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
```

```

OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN SEQUENCE FROM N.A.
RA Penella L.O.F., Ruiz M.F., Granadino B., Vicente L., Ortega A.,
RA Valcarlos J., Sanchez L.;
RA Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RL EMBL; AJ243599; CAB46637.1;
DR Flybase; FBgn0000662; fl(2)d.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0019099; P:female germ-line sex determination; NAS.
DR GO; GO:0007539; P:primary sex determination, soma; NAS.
DR GO; GO:0000391; P:regulation of alternative nuclear mRNA splicing; IPI.
DR GO; GO:0007530; P:sex determination; IPI.
DR InterPro; IPR009053; Pfam01d.
SQ SEQUENCE 539 AA; 59608 MW; FB20579F83CFCD5 CRC64;

Query Match          32.9%; Score 528.5; DB 2; Length 539;
Best Local Similarity 33.9%; Pred. No. 5.8e-19;
Matches 127; Conservative 62; Mismatches 113; Indels 73; Gaps 8;

QY 2 RIATKGQDIQECTQIQLKQVQPSVAQLRSTWVDPAINLFFLKMKGLEOTKDKLEQA 61
Db 145 RLANKQEFQDVQSYQIAEYKQAQPTALATLALDPAVNLFFELKKELKATKALEET 204
QY 62 QNELSAWFTDSTQTKLMAKRMLOENQELGRLSQGRIAOELAEALQKYSBELK 121
Db 205 QNELSAWFTDSTQTKLMAKRMLOENQELGRLSQGRIAOELAEALQKYSBELK 121
QY 122 SSQDELNDFFIQLDVEVGMSSTLVLQQLKQETROQLAQVQQQSQ-----ASA 171
Db 265 KSQSELDFLOLDEDDVEGMSSTLFLQQLKQETROQLAQVQQQSQ-----ASA 171
QY 172 PSTSRRTASEPVEQSEATSKDCSLKTNQPSNGSSRQRTSGSGFHRGENTEDDFPS-- 229
Db 325 AATNGGT-NTTINKLETIHEDAC-MANNPTNPD-----YNGNTNNEQIAAAPQ 371
QY 230 --PGNGKNSNSSEERTGGGGVYNQLSAGYSDVSDTGSNSL-----272
Db 372 IPPSDGSGNMGNAARLARKNYQEEALFTVVVPTPTPGVNNVQGAAPIREVTAPRTL 431
QY 273 -----THOSNPTDSSHDPOE-----EKAVSGKGNRTVGSRHV-----304
Db 432 PPKSKLGRITTRNSQLEEDHQVPTTPVAPMVDVNAVAGMASEEAAAAAANNNTG 491
QY 305 ---QNGLDSSVNVQV 316
Db 492 IIPETGVQGVPEV 506

RESULT 9
Q8ML74 PRELIMINARY; PRT; 412 AA.
ID Q8ML74
AC Q8ML74
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE CG315-PB.
GN Name=fl(2)d; ORFNames=CG6315;
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN SEQUENCE FROM N.A.
RA MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,

```

```

RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkuch C., Baldwin D.,
RA Balaw R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahike C., Davenport L.B., Davies P.,
RA De Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.B., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush P., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinart K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kianos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Turner R., Venter E., Wang A.H., Wang X.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang Q., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
RA Pacle J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.,
RT "Finishing a whole-genome shotgun: release 3 of the Drosophila
RL melanogaster euchromatic genome sequence."
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RL a genomics perspective."
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RL systematic review."
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN [5]
RP SEQUENCE FROM N.A.
RG FLYBASE;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RG FLYBASE;

```


RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celniker S.,
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF003817; AAF58334.1; --
 DR EMBL; AY069478; AAL39623.1; --
 DR IntAct; Q9V6T8; --
 DR FlyBase; FBgn000662; fl(2)d.
 DR CO; GO:0005634; C:nucleus; IDA.
 DR CO; GO:0019599; P:female germ-line sex determination; NAS.
 DR CO; GO:0007539; P:primary sex determination, soma; NAS.
 DR CO; GO:0000381; P:regulation of alternative nuclear mRNA splicing; IPI.
 DR CO; GO:0007530; P:sex determination; IPI.
 SQ SEQUENCE 536 AA; 59268 MW; 732DF06E999ECBEF CRC64;

Query Match 32.7%; Score 525.5; DB 2; Length 536;
 Best Local Similarity 33.9%; Pred. No. 8.1e-19;
 Matches 127; Conservative 61; Mismatches 114; Indels 73; Gaps 8;

QY 2 RLATKGDIQECTTQIQLKQVQPSVAQLRSTWDPAINLFFLKMGGELEOTKQKLEQA 61
 DB 142 RLATKEQFQYVSQIAYKAAQAPTALALRTALDPAVNLLFRLKKELKATKALEET 201
 QY 62 QNELSAWKTTPDSOTGKLMACRMILIQENQELGRQLSQRIAGLEALALQKYSSEELK 121
 DB 202 QNELSAWKTTPDSNTGKLMACRMILYQENELGKMTSNGRLAKLETALQKSFSEVK 261
 QY 122 SSQDELNDFFIQLDEVEGCMSTILVLOQLKETRQQLAQYQQQSQ-----ASA 171
 DB 262 KSQELDDFLQELDEDEVEGCMSTILVLOQLKETRQQLAQYQQQSQ-----ASA 321
 QY 172 PSTRTTASEVEQSEATSKDCSLRTNGSPNSGSSRQRTSGSGFHREGNTTDDFPSSP- 230
 DB 322 AATNGGT-NTTINKLETIHEDAC-WANNPTNPD-----YNGTNNQIAAVPQ 368
 QY 231 ---GNGKSSNSERTGRGGSGYVNSLQAGYVSDPTSGNSL----- 272
 DB 369 IPLSDGSGNMGNAARLARKENYQEEALPTVVVPTFTFVGNVQAPPIREVTAPRTL 428
 QY 273 -----THOSNTDSSHDPOE-----EKAVSGKGNRTVGRHV----- 304
 DB 429 PPKSKLGLITRNSQLDEHQVPTTFAVPMIVDNAVAGMASEEAAAAAANNSTG 488
 QY 305 ---ONGLDSSVNVVG 316
 DB 489 IIPETGVGVFVFG 503

RESULT 11
 QYQ0B0 PRELIMINARY; PRT; 272 AA.
 AC QYQ0B0;
 DT 01-MAR-2004 (TrEMBLrel. 26, Created)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE ACP29790 (Fragment)
 GN Name=agCS3978; ORFNames=ENSG000000009104;
 OS Anopheles gambiae str. PEST.
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
 OX NCBI_TaxID=180454;
 RN [1]
 RC STRAIN=PEST;
 RA Anopheles Genome Sequencing Consortium;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 CC -!- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 DR EMBL; AAB01008986; EAA00485.1; --
 FT NON_TER 1
 FT NON_TER 272
 SQ SEQUENCE 272 AA; 30286 MW; 81C20E25C072EE28 CRC64;

Query Match 32.2%; Score 518; DB 2; Length 272;
 Best Local Similarity 39.2%; Pred. No. 9.1e-19;
 Matches 120; Conservative 48; Mismatches 72; Indels 66; Gaps 6;

QY 2 RLATKGDIQECTTQIQLKQVQPSVAQLRSTWDPAINLFFLKMGGELEOTKQKLEQA 61
 DB 30 RVAVQEQMRDFAKQIAELKSAQAPGPAALRSALLDPAVNILFQKLKAELOKAKLEET 89
 QY 62 QNELSAWKTTPDSOTGKLMACRMILIQENQELGRQLSQRIAGLEALALQKYSSEELK 121
 DB 90 QNELSAWKTTPDSNTGKLMACRMILYQENELGKMTSNGRLAKLESEALQKSTNEEVK 149
 QY 122 SSQDELNDFFIQLDEVEGCMSTILVLOQLKETRQQLAQYQQQSQASAPSTRTTASE 181
 DB 150 KSQELDDFLQELDEDEVEGCMQCTIVFLQQLAKQTKDARNEIEKVIQLRAYIAANL--- 205
 QY 182 PVEQSEATSKDCSLRTNGSPNSGSSRQRTSGSGFHREGNTTDDFPSSPGNKSNSSE 241
 DB 206 -----GPFSGS-----GN-----NPTSGSTPTD 223
 QY 242 ERTGRGGGYVNSLQAGYVSDPTSGNSLTHQSDTSSHDPOEKAVSGKGNRTVGS 301
 DB 224 ERS-----SSMMWMDVQDSNSNSH---GHDAYHP-----DGGGPDSTIGG 263
 QY 302 RHVQNG 307
 DB 264 RGVRLG 269

RESULT 12
 QYQZS4 PRELIMINARY; PRT; 74 AA.
 AC QYQZS4;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE PNAS-132.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Yu W.-O., Sun B.-Z., Chai Y.-B., Zhu F., Liu X.-S., Li Z., Lu F.,
 RA Yan W., Yang H., Zhao Z.-L.,
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF271190; AAK07548.1; --
 SQ SEQUENCE 74 AA; 8690 MW; 51DC8BE6CF09B0D9 CRC64;

Query Match 22.2%; Score 357; DB 2; Length 74;
 Best Local Similarity 95.9%; Pred. No. 2.7e-11;
 Matches 70; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLATKGDIQECTTQIQLKQVQPSVAQLRSTWDPAINLFFLKMGGELEOTKQKLEQ 60
 DB 1 RLATKEQEMOECTTQIQLKQVQPSVAQLRSTWDPAINLFFLKMGGELEOTKQKLEQ 60
 QY 61 AQNELSAWKTTPD 73
 DB 61 AQNELSAWKTTPD 73

RESULT 13
 QYQZS8 PRELIMINARY; PRT; 330 AA.
 AC QYQZS8;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE Putative FKBP12 interacting protein FIP37 (FIP37).
 GN Name=FIP37; Synonyms=At3g54170, F24B22.130;
 OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=99024949; PubMed=9807817;
RP Faure J.-D., Gingerich D., Howell S.H.;
RT "An Arabidopsis immunophilin, AtFKBP12, binds to AtPIP37 (FKBP
interacting protein) in an interaction that is disrupted by FK506.";
RL Plant J. 15:783-789 (1998).
RN [2]
RN SEQUENCE FROM N.A.
RP Faure J.-D., Gingerich D., Howell S.H.;
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RN SEQUENCE FROM N.A.
RP Bloeker H., Mewes H.W., Lemcke K., Mayer K.F.X., Quetier F.,
RA Salanoubat M.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RN SEQUENCE FROM N.A.
RP EU Arabidopsis sequencing project;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RN SEQUENCE FROM N.A.
RP Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
RA Seki M., Shinn P., Southwick A., Shinzaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
RN [6]
RN SEQUENCE FROM N.A.
RP Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,
RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,
RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
RA Seki M., Shinn P., Southwick A., Shinzaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
RN [7]
RN EMBL; AF084570; AAC72922.1; -;
DR EMBL; AL132957; CAB70991.1; -;
DR EMBL; AY063940; AAL36296.1; -;
DR EMBL; AY113943; AAM44991.1; -;
DR PIR; T47576; T47576.
SQ SEQUENCE 330 AA; 37214 MW; 6B4070DSAD332004 CRC64;
Query Match 17.5%; Score 282; DB 2; Length 330;
Best Local Similarity 33.5%; Pred. No. 8.4e-07;
Matches 55; Conservative 51; Mismatches 58; Indels 0; Gaps 0;
QY 1 MRLATKGDIOECTTQIQLKQVQPSVAQLRSTWDPAINLFLKMKGELEOTKDKLEQ 60
DB 136 VQYAKREQEMAEKSAVDLKSQKLPASMQARRLLDPAIHEEFSRLNVLVEKKXIKE 195
QY 61 AQNELSAWKFTPDSTQTKMLMAKCRMLIQENQELGRQLSQGRIAGLEALQKYSLEL 120
DB 196 LQDNIAAVTFTFQSKNGKGLMAKCRMLTQEEENEIGHQAAGKIHLEAIKLAQKSNAL 255
QY 121 KSSQDELNDFTIQLDEEVEGQSTILVLQOOLKETROOLAQYQ 164
DB 256 RSQFEGLYKHMEELNDVRSNETVILIQLEEKKEIEIRVKK 299
RESULT 14
Q6PNZ9 PRELIMINARY; PRT; 267 AA.
ID Q6PNZ9
AC Q6PNZ9;

DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE FKBP12-like protein (Fragment).
OS Mirabilis jalapa (Garden four-o'clock).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllales; Nyctaginaceae; Mirabilis.
OX NCBI_TaxID=3538;
RN [1]
RN SEQUENCE FROM N.A.
RC TISSUE=Petiole abscission zone;
RA Meir S., Hunter D.A., Chen J.-C., Reid M.S.;
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY589703; AAT07458.1; -;
FT NON_TER 1
FT NON_TER 267
SQ SEQUENCE 267 AA; 30178 MW; D2A198B6C3A3A4AE CRC64;
Query Match 13.6%; Score 218; DB 2; Length 267;
Best Local Similarity 37.3%; Pred. No. 0.0011;
Matches 44; Conservative 32; Mismatches 42; Indels 0; Gaps 0;
QY 4 ATKGDIOECTTQIQLKQVQPSVAQLRSTWDPAINLFLKMKGELEOTKDKLEQ 63
DB 150 AKRQEIAELKAAIRELSQLKPLSMQTRRLDPAIHEEFTLKNLVEEKVKVQLQD 209
QY 64 ELSAWKFTPDSTQTKMLMAKCRMLIQENQELGRQLSQGRIAGLEALQKYSLEL 121
DB 210 NINAVNFTANSKMGKMLMAKCKTLOEENEIGHQAAGKIHLEAIKLAQKSNSEIR 267
RESULT 15
AAT07458 PRELIMINARY; PRT; 267 AA.
ID AAT07458
AC AAT07458;
DT 12-MAY-2004 (TrEMBLrel. 27, Created)
DT 12-MAY-2004 (TrEMBLrel. 27, Last sequence update)
DE FKBP12-like protein (Fragment).
OS Mirabilis jalapa (Garden four-o'clock).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllales; Nyctaginaceae; Mirabilis.
OX NCBI_TaxID=3538;
RN [1]
RN SEQUENCE FROM N.A.
RC TISSUE=Petiole abscission zone;
RA Meir S., Hunter D.A., Chen J.-C., Reid M.S.;
RL "Gene changes associated with the acquisition of abscission zone
competence in Mirabilis jalapa.";
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY589703; AAT07458.1; -;
FT NON_TER 1
FT NON_TER 267
SQ SEQUENCE 267 AA; 30178 MW; D2A198B6C3A3A4AE CRC64;
Query Match 13.6%; Score 218; DB 2; Length 267;
Best Local Similarity 37.3%; Pred. No. 0.0011;
Matches 44; Conservative 32; Mismatches 42; Indels 0; Gaps 0;
QY 4 ATKGDIOECTTQIQLKQVQPSVAQLRSTWDPAINLFLKMKGELEOTKDKLEQ 63
DB 150 AKRQEIAELKAAIRELSQLKPLSMQTRRLDPAIHEEFTLKNLVEEKVKVQLQD 209
QY 64 ELSAWKFTPDSTQTKMLMAKCRMLIQENQELGRQLSQGRIAGLEALQKYSLEL 121
DB 210 NINAVNFTANSKMGKMLMAKCKTLOEENEIGHQAAGKIHLEAIKLAQKSNSEIR 267
Search completed: November 14, 2004, 12:36:30
Job time : 194 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2004, 12:14:48 ; Search time 156 Seconds
(without alignments)
733.555 Million cell updates/sec

Title: US-10-030-389-2
Perfect score: 1608
Sequence: 1 MRLATKGQDIQECTTQIQL.....GSRHVQNGLDSSVNVQGSVL 319

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Genesep23Sep04:*

- 1: genesep1990s:*
- 2: genesep1990s:*
- 3: genesep2000s:*
- 4: genesep2001s:*
- 5: genesep2002s:*
- 6: genesep2003as:*
- 7: genesep2003bs:*
- 8: genesep2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1608	100.0	319	4	AAB67048 Human imm
2	1570	97.6	388	6	ABO53045 Human put
3	525.5	32.7	536	4	ABB62262 Drosophil
4	405.	25.2	671	4	ABG20586 Novel hum
5	357	22.2	151	6	ABO53092 Human put
6	357	22.2	151	8	ADO58680 Human reg
7	282	17.5	278	3	AAQ41464 Arabidops
8	282	17.5	330	3	AAQ41463 Arabidops
9	271	16.9	186	3	AAQ41465 Arabidops
10	256	15.9	352	8	ADM48175 Polypepti
11	178	11.1	102	4	ABG10314 Novel hum
12	166	10.3	524	6	ABU29734 Protein e
13	166	10.3	525	7	ABU29734 Protein e
14	156.5	9.7	511	5	ABG93255 C. albica
15	156	9.7	897	2	AAW2751 Murine EG
16	156	9.7	897	2	AAW47118 Murine EG
17	156	9.7	897	2	AAW94406 Murine ep
18	152.5	9.5	1031	4	AAU35339 Enterococ
19	150	9.3	431	7	ADD93650 Streptoco
20	150	9.3	456	5	ABE55584 Lactococ
21	149	9.3	431	7	ADD93653 Streptoco
22	147	9.1	912	4	ABE60403 Drosophil
23	147	9.1	2112	4	ABE60403 Drosophil
24	146	9.1	431	7	ADD93649 Streptoco
25	145	9.0	3502	4	ABB58382 Drosophil

26	144.5	9.0	877	8	ADN72429 Thale cre
27	144	9.0	461	2	AAR14530 Usp45 pro
28	144	9.0	461	2	AAR14530 MSP encod
29	144	9.0	716	2	AY22176 Drosophil
30	144	9.0	1807	8	ADO17160 Novel exp
31	142.5	8.9	744	4	ABU29451 Protein e
32	142.5	8.9	1061	6	ABU29451 Protein e
33	141.5	8.8	432	7	ADD93651 Streptoco
34	141.5	8.8	562	4	ABB60336 Drosophil
35	141	8.8	294	5	ABP41663 Human ova
36	141	8.8	1802	3	ABP41663 Human ova
37	141	8.8	1802	3	ABP41663 Human ova
38	141	8.8	3248	2	AAAR9795 Cell wall
39	140	8.7	108	4	ABG20587 Kinetoch
40	140	8.7	1031	7	ADB67093 Kinesin h
41	140	8.7	1032	8	ADL99364 Nanos truc
42	140	8.7	1091	4	ABB68998 Drosophil
43	140	8.7	1920	5	ABG97508 Human NOV
44	139.5	9.7	1135	3	AAAY68784 Amino aci
45	139	8.6	1017	7	ADJ68352 Human hea

ALIGNMENTS

RESULT 1
AAB67048
ID AAB67048 standard; protein; 319 AA.
XX
AC AAB67048;
XX
DT 09-APR-2001 (first entry)
XX
DE Human immune response molecule (IMUN) protein SEQ ID NO: 2.
XX
KW Human; IMUN; immune response molecule; autoimmune disorder;
KW inflammatory disorder; cell proliferation disorder; cancer.
XX
OS Homo sapiens.
XX
PN WO200102569-A2.
XX
PD 11-JAN-2001.
XX
PF 06-JUL-2000; 2000WO-US018505.
XX
PR 06-JUL-1999; 99US-0142572P.
PR 09-SEP-1999; 99US-0153170P.
XX
(INCY-) INCYTE GENOMICS INC.
PA Tang YT, Yue H, Yang J, Azimzai Y, Baughn MR, Lu DAM;
PI WPI; 2001-123113/13.
DR N-PSDB; AAF31245.
XX
PT Novel human immune response molecules (IMUN), useful for diagnosing,
PT treating and preventing disorders associated with abnormal expression of
PT IMUN, e.g. Addison's disease, allergies, anemia, asthma and
PT atherosclerosis.
XX
PS Claim 2; Page 82; 95pp; English.
XX
CC The present invention provides the protein and coding sequences for ten
CC human immune response molecules (designated IMUN-1 to IMUN-10). The
CC sequences can be used in the diagnosis, treatment and prevention of
CC autoimmune and inflammatory disorders such as AIDS, atherosclerosis,
CC asthma, allergies, Crohn's disease, multiple sclerosis, irritable bowel
CC syndrome, psoriasis, rheumatoid arthritis and infections, and cell
CC proliferation disorders including arteriosclerosis, cirrhosis and cancer
XX
SQ Sequence 319 AA;

Query Match 100.0%; Score 1608; DB 4; Length 319;
 Best Local Similarity 100.0%; Pred. No. 5.7e-121; Indels 0; Gaps 0;
 Matches 319; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRLATKGQDIQECTTQIQYLVKQVQPSVAQLRSTWVDPAINLFFFLKMGELQTKDKLEQ 60
 DB 1 MRLATKGQDIQECTTQIQYLVKQVQPSVAQLRSTWVDPAINLFFFLKMGELQTKDKLEQ 60

QY 61 AQNELSAWKFTPDSTQTKMLAKCMLIQENQELGRQLSQGRIQLAEALQKYSSEL 120
 DB 61 AQNELSAWKFTPDSTQTKMLAKCMLIQENQELGRQLSQGRIQLAEALQKYSSEL 120

QY 121 KSSQDELNDFFIQLDEEVEGQSTILVQLQQLKTRQQLAQYQQQSQASAPSTRITAS 180
 DB 121 KSSQDELNDFFIQLDEEVEGQSTILVQLQQLKTRQQLAQYQQQSQASAPSTRITAS 180

QY 181 EPVEQSEATSKDCSRLTNGPNSGSSRRQRTSGSGFHREGNTTDDFPSSPGNGKSNSS 240
 DB 181 EPVEQSEATSKDCSRLTNGPNSGSSRRQRTSGSGFHREGNTTDDFPSSPGNGKSNSS 240

QY 241 EERTGRGGGYVNLQSLAGYVESVDSPGSENLSLTHQSNLTDSSHPDQBEKAVSGKGNRTVG 300
 DB 241 EERTGRGGGYVNLQSLAGYVESVDSPGSENLSLTHQSNLTDSSHPDQBEKAVSGKGNRTVG 300

QY 301 SRHVQGLDSSVNVQGSVL 319
 DB 301 SRHVQGLDSSVNVQGSVL 319

RESULT 2
 ABO53045
 ID ABO53045 standard; protein; 388 AA.
 AC ABO53045;
 XX
 DT 10-OCT-2003 (first entry)
 DE Human putative spliceosome associated protein (SAP) #21.
 XX
 KW Human; SAP; spliceosome associated protein; ribonucleoprotein;
 KW RNP complex; RNA affinity substrate; RNP assembly sequence;
 KW spliceosomal complex; hnRNP complex; mRNA export complex;
 KW mRNA localisation complex; RNA editing complex; intron complex;
 KW H complex; telomerase complex; fragile X protein complex;
 KW reverse transcriptase complex; gene splicing complex.
 XX
 OS Homo sapiens.
 XX
 FN US2003068803-A1.
 XX
 PD 10-APR-2003.
 XX
 PF 14-JAN-2002; 2002US-00047991.
 XX
 PR 12-JAN-2001; 2001US-0261521P.
 XX
 PA (REED/) REED R.
 PA (ZHOU/) ZHOU Z.
 XX
 PI Reed R, Zhou Z;
 XX
 DR WPI; 2003-540885/51.
 XX
 PT Isolating ribonucleoprotein complex, by contacting RNA affinity substrate
 PT having ribonucleoprotein assembly sequence and affinity tag, with protein
 PT mixture, subjecting complex formed to chromatography, affinity selection.
 XX
 PS Claim 24; Page; 39pp; English.
 XX
 CC The invention relates to forming (M1) an isolated ribonucleoprotein (RNP)
 CC complex (C), involves contacting an RNA affinity substrate (S) comprising
 CC an RNP assembly sequence (AS) and an affinity tag, with a protein mixture
 CC to permit formation of (C) on AS, subjecting (C) to chromatographic

CC separation, and subjecting (C) to affinity selection, where the affinity
 CC tag (e.g. bacteriophage MS2 coat protein in a fusion protein with E. coli
 CC maltose binding protein) binds to an affinity matrix. Also included are
 CC an isolated spliceosome preparation (isolated by (M1)), a RNA comprising
 CC an RNP complex binding site and at least one phase coat protein
 CC recognition site, a nucleic acid encoding the RNA, and treating (M2) a
 CC subject having a disorder associated with abnormal RNP complexes (by
 CC obtaining a sample of cells from a subject, purifying RNP complexes from
 CC the cells of the subject by (M1), determining the presence in the
 CC purified RNP complexes of one or more proteins, and normalising the
 CC amount of RNPs in the subject. (M1) is useful for forming an isolated RNP
 CC complex selected from a spliceosomal complex (selected from E, A, B and C
 CC complex), an hnRNP complex, an mRNA export complex, an mRNA localisation
 CC complex, an RNA editing complex, an intron complex, or an H complex. (M1)
 CC is useful in a diagnostic assay for determining whether a subject has
 CC abnormal RNP complexes, (M2) is useful for treating a subject having a
 CC disorder associated with abnormal RNP complexes. (M1) is useful for
 CC forming an isolated RNP complex such as a telomerase complex, a fragile X
 CC protein complex, a reverse transcriptase complex or a gene splicing
 CC complex. The present sequence represents a putative novel human
 CC spliceosome associated protein (SAP) isolated by the methods of the
 CC invention. Note: The present sequence is not shown in the specification
 CC but was obtained from Genbank or Swissprot using the information provided
 CC in table 2 of the specification
 XX
 SQ Sequence 388 AA;

Query Match 97.6%; Score 1570; DB 6; Length 388;
 Best Local Similarity 99.0%; Pred. No. 8.6e-118;
 Matches 311; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MRLATKGQDIQECTTQIQYLVKQVQPSVAQLRSTWVDPAINLFFFLKMGELQTKDKLEQ 60
 DB 75 MRLATKEQEMQECTTQIQYLVKQVQPSVAQLRSTWVDPAINLFFFLKMGELQTKDKLEQ 134

QY 61 AQNELSAWKFTPDSTQTKMLAKCMLIQENQELGRQLSQGRIQLAEALQKYSSEL 120
 DB 135 AQNELSAWKFTPDSTQTKMLAKCMLIQENQELGRQLSQGRIQLAEALQKYSSEL 194

QY 121 KSSQDELNDFFIQLDEEVEGQSTILVQLQQLKTRQQLAQYQQQSQASAPSTRITAS 180
 DB 195 KSSQDELNDFFIQLDEEVEGQSTILVQLQQLKTRQQLAQYQQQSQASAPSTRITAS 254

QY 181 EPVEQSEATSKDCSRLTNGPNSGSSRRQRTSGSGFHREGNTTDDFPSSPGNGKSNSS 240
 DB 255 EPVEQSEATSKDCSRLTNGPNSGSSRRQRTSGSGFHREGNTTDDFPSSPGNGKSNSS 314

QY 241 EERTGRGGGYVNLQSLAGYVESVDSPGSENLSLTHQSNLTDSSHPDQBEKAVSGKGNRTVG 300
 DB 315 EERTGRGGGYVNLQSLAGYVESVDSPGSENLSLTHQSNLTDSSHPDQBEKAVSGKGNRTVG 374

QY 301 SRHVQGLDSSVNV 314
 DB 375 SRHVQGLDSSVNV 388

RESULT 3
 ABB62262
 ID ABB62262 standard; protein; 536 AA.
 AC ABB62262;
 XX
 DT 26-MAR-2002 (first entry)
 DE Drosophila melanogaster polypeptide SEQ ID NO 13578.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 CC
 FN WO200171042-A2.
 XX

PD 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US009231.
XX 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX (PEKE) PE CORP NY.
PA Venter JC, Adams M, Li PWD, Myers EW;
XX WPI: 2001-656860/75.
XX N-PSDB; ABL06365.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX Disclosure; SEQ ID NO 13578; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signaling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL15175) and the encoded proteins (AB857737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 536 AA;
XX
Query Match 32.7%; Score 525.5; DB 4; Length 536;
Best Local Similarity 33.9%; Pred. No. 1.8e-33;
Matches 127; Conservative 61; Mismatches 114; Indels 73; Gaps 8;
QY 2 RLATGQDIQCTQIQYLVKQVQPSVAQLRSTWVDPAINLFFLMKGELEQTKLEQA 61
DB 142 FLANKEQFQDYVSQIAEYKQQAFTALATLTDPAVNLFFELKKELKATKLEET 201
QY 62 QNELSAWFTPDSTGKLMKAKRMLOENQELGRLSQRIAGLEALALOKKYSLEK 121
DB 202 QNELSAWFTPDSTGKLMKACRLLYQENELGKMTSNGRLAKLETALAMQKSFSEVK 261
QY 122 SSOELNDFIQLDEVEGMOSTILVLOQLKETROQLAQYQQOQSO-----ASA 171
DB 252 KSQSELDLDFLELDEDEVEGMOSTILVLOQLKETROQLAQYQQOQSO----- 321
QY 172 PSTSTTASPEVQSEATSKDCSLTNGPSNGSSSRQRTSGSGFHREGNTTDEDFPSP- 230
DB 322 AATNGGT-NTTINKLETHEDAC-MANNPTNPD-----YNGNTNNEQIAAVPQ 368
QY 231 ---GNGKSSNSSEERTGSGGYVQLSAGVESVDSPTGSENSI----- 272
DB 369 IPLSDDGNGMNGAARLARKNYQEEALPTVVVPTPTVGNVQEAAPPREVTAPRTL 428
QY 273 -----THQNDTSDSHDPOE-----EKAVSGKGNRTVGSRHV----- 304
DB 429 PPKSKLRLGITTRNSQLEEDHQPTTTPVAVPMIVDNVAGVASEAAAAAANNNTG 488
QY 305 ---QNGLSSVNVQ 316
DB 489 IIPETGVQGVPEV 503
RESULT 4
ABG20586
ID ABG20586 standard; protein; 671 AA.
XX
AC ABG20586;
XX
DT 13-FEB-2002 (first entry)

XX Novel human diagnostic protein #20577.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US008631.
XX 31-MAR-2000; 2000US-00540217.
XX 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI: 2001-639362/73.
XX N-PSDB; AAS64773.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX Claim 20; SEQ ID NO 50945; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have application in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX Sequence 671 AA;
XX
Query Match 25.2%; Score 405; DB 4; Length 671;
Best Local Similarity 63.2%; Pred. No. 1.2e-23;
Matches 98; Conservative 2; Mismatches 23; Indels 32; Gaps 5;
QY 126 ELNDFIQLDEVEGMOSTILVLOQLKETROQLAQYQQOQSOASAPSTRTTASEPVEQ 195
DB 381 ELNDFIQLDEVEGMOSTILVLOQLKETROQLAQYQQOQSOASAPSTRTTASEPVEQ 440
QY 186 SEATSKDCSLTNGPSNGSS-----SRQTSKSGFPHREG-----NTTEDD 225
DB 441 SEATSKDCSLTNGPSYGSFPRGRSHSPAAPRGRKEGREAVALGARGGRTIMEQL 500
QY 226 PPSG-----PG-----NGKSSNS--EERTGRGG 248
DB 501 LPGNTPFPFGAWRELGRPESSASSVREAAEGSG 535
RESULT 5

[illegible]

PR	07-JUN-1999;	99US-0137724P.	PR	06-AUG-1999;	99US-0147416P.
PR	08-JUN-1999;	99US-0138094P.	PR	09-AUG-1999;	99US-0147493P.
PR	10-JUN-1999;	99US-0138540P.	PR	09-AUG-1999;	99US-0147935P.
PR	10-JUN-1999;	99US-0138847P.	PR	10-AUG-1999;	99US-0148171P.
PR	14-JUN-1999;	99US-0139119P.	PR	11-AUG-1999;	99US-0148319P.
PR	16-JUN-1999;	99US-0139452P.	PR	12-AUG-1999;	99US-0148341P.
PR	16-JUN-1999;	99US-0139453P.	PR	13-AUG-1999;	99US-0148565P.
PR	17-JUN-1999;	99US-0139492P.	PR	13-AUG-1999;	99US-0148684P.
PR	18-JUN-1999;	99US-0139454P.	PR	16-AUG-1999;	99US-0149368P.
PR	18-JUN-1999;	99US-0139455P.	PR	17-AUG-1999;	99US-0149175P.
PR	18-JUN-1999;	99US-0139456P.	PR	18-AUG-1999;	99US-0149426P.
PR	18-JUN-1999;	99US-0139457P.	PR	20-AUG-1999;	99US-0149722P.
PR	18-JUN-1999;	99US-0139458P.	PR	20-AUG-1999;	99US-0149723P.
PR	18-JUN-1999;	99US-0139459P.	PR	20-AUG-1999;	99US-0149929P.
PR	18-JUN-1999;	99US-0139460P.	PR	23-AUG-1999;	99US-0149902P.
PR	18-JUN-1999;	99US-0139461P.	PR	23-AUG-1999;	99US-0149930P.
PR	18-JUN-1999;	99US-0139462P.	PR	25-AUG-1999;	99US-0150566P.
PR	18-JUN-1999;	99US-0139463P.	PR	26-AUG-1999;	99US-0150884P.
PR	18-JUN-1999;	99US-0139750P.	PR	27-AUG-1999;	99US-0151065P.
PR	21-JUN-1999;	99US-0139817P.	PR	27-AUG-1999;	99US-0151066P.
PR	22-JUN-1999;	99US-0139899P.	PR	27-AUG-1999;	99US-0151080P.
PR	23-JUN-1999;	99US-0140333P.	PR	27-AUG-1999;	99US-0151080P.
PR	23-JUN-1999;	99US-0140334P.	PR	30-AUG-1999;	99US-0151303P.
PR	24-JUN-1999;	99US-0140695P.	PR	31-AUG-1999;	99US-0151438P.
PR	28-JUN-1999;	99US-0140823P.	PR	01-SEP-1999;	99US-0151930P.
PR	29-JUN-1999;	99US-0140991P.	PR	07-SEP-1999;	99US-0152363P.
PR	30-JUN-1999;	99US-0141287P.	PR	10-SEP-1999;	99US-0153070P.
PR	01-JUL-1999;	99US-0141842P.	PR	13-SEP-1999;	99US-0153758P.
PR	01-JUL-1999;	99US-0142154P.	PR	15-SEP-1999;	99US-0154018P.
PR	02-JUL-1999;	99US-0142055P.	PR	16-SEP-1999;	99US-0154039P.
PR	06-JUL-1999;	99US-0142330P.	PR	20-SEP-1999;	99US-0154779P.
PR	08-JUL-1999;	99US-0142803P.	PR	22-SEP-1999;	99US-0155139P.
PR	09-JUL-1999;	99US-0142920P.	PR	23-SEP-1999;	99US-0155486P.
PR	12-JUL-1999;	99US-0142977P.	PR	24-SEP-1999;	99US-0155659P.
PR	13-JUL-1999;	99US-0143542P.	PR	28-SEP-1999;	99US-0156458P.
PR	14-JUL-1999;	99US-0143624P.	PR	29-SEP-1999;	99US-0156596P.
PR	15-JUL-1999;	99US-0144005P.	PR	04-OCT-1999;	99US-0157117P.
PR	16-JUL-1999;	99US-0144085P.	PR	05-OCT-1999;	99US-0157753P.
PR	16-JUL-1999;	99US-0144086P.	PR	06-OCT-1999;	99US-0157865P.
PR	19-JUL-1999;	99US-0144332P.	PR	07-OCT-1999;	99US-0158029P.
PR	19-JUL-1999;	99US-0144331P.	PR	08-OCT-1999;	99US-0158232P.
PR	19-JUL-1999;	99US-0144332P.	PR	12-OCT-1999;	99US-0158369P.
PR	19-JUL-1999;	99US-0144333P.	PR	13-OCT-1999;	99US-0159293P.
PR	19-JUL-1999;	99US-0144334P.	PR	13-OCT-1999;	99US-0159294P.
PR	19-JUL-1999;	99US-0144335P.	PR	13-OCT-1999;	99US-0159295P.
PR	20-JUL-1999;	99US-0144332P.	PR	14-OCT-1999;	99US-0159329P.
PR	20-JUL-1999;	99US-0144332P.	PR	14-OCT-1999;	99US-0159330P.
PR	20-JUL-1999;	99US-0144882P.	PR	14-OCT-1999;	99US-0159331P.
PR	21-JUL-1999;	99US-0144814P.	PR	14-OCT-1999;	99US-0159637P.
PR	21-JUL-1999;	99US-0145086P.	PR	18-OCT-1999;	99US-0159584P.
PR	21-JUL-1999;	99US-0145088P.	PR	21-OCT-1999;	99US-0160741P.
PR	22-JUL-1999;	99US-0145085P.	PR	21-OCT-1999;	99US-0160767P.
PR	22-JUL-1999;	99US-0145087P.	PR	21-OCT-1999;	99US-0160768P.
PR	22-JUL-1999;	99US-0145089P.	PR	21-OCT-1999;	99US-0160770P.
PR	22-JUL-1999;	99US-0145129P.	PR	21-OCT-1999;	99US-0160814P.
PR	23-JUL-1999;	99US-0145218P.	PR	21-OCT-1999;	99US-0160815P.
PR	23-JUL-1999;	99US-0145224P.	PR	22-OCT-1999;	99US-0160980P.
PR	26-JUL-1999;	99US-0145276P.	PR	22-OCT-1999;	99US-0160981P.
PR	27-JUL-1999;	99US-0145913P.	PR	22-OCT-1999;	99US-0160989P.
PR	27-JUL-1999;	99US-0145918P.	PR	23-OCT-1999;	99US-0161404P.
PR	27-JUL-1999;	99US-0145919P.	PR	23-OCT-1999;	99US-0161405P.
PR	28-JUL-1999;	99US-0145951P.	PR	25-OCT-1999;	99US-0161406P.
PR	02-AUG-1999;	99US-0146386P.	PR	26-OCT-1999;	99US-0161359P.
PR	02-AUG-1999;	99US-0146388P.	PR	26-OCT-1999;	99US-0161360P.
PR	02-AUG-1999;	99US-0146389P.	PR	26-OCT-1999;	99US-0161361P.
PR	03-AUG-1999;	99US-0147038P.	PR	28-OCT-1999;	99US-0161920P.
PR	04-AUG-1999;	99US-0147204P.	PR	28-OCT-1999;	99US-0161922P.
PR	04-AUG-1999;	99US-0147302P.	PR	28-OCT-1999;	99US-0161933P.
PR	05-AUG-1999;	99US-0147192P.	PR	29-OCT-1999;	99US-0162142P.
PR	05-AUG-1999;	99US-0147260P.			
PR	06-AUG-1999;	99US-0147303P.			

Query Match 17.5%; Score 282; DB 3; Length 330;
Best Local Similarity 33.5%; Pred. No. 3.7e-14;
Matches 55; Conservative 51; Mismatches 58; Indels 0; Caps 0;

```

QY      1  MRLATKGQDIQCTTQIQYLVKQVQSPVAQRSTNVDPAINLFFLKMKGELEQTKDKLEQ 60
DB      136  VYAKREQEMAEKLSAVRDLKSQLPASMQARRLLDPAIHEEFSRLKNLVEEKDKKIKE 195
QY      61  AQNELSAWKFTPDSTQTKKMAKCRMLIQENQELGRQLSQGRIQAQLEALALOKKYSEEL 120
DB      196  LDNIAAATVFPQSKNGKMLAKCETLQEEHEEIGHQAEGKIHELAKLAWQKSQNAEL 255
QY      121  KSSQDELNDFIQLDEEVGQSTILVLIQQQLKETROQLAQYQ 164
DB      256  RSQFEGLYKHMBELTNDVRSNETVIIQEXLEKEKEIERVVK 299

RESULT 9
AAG41465
ID  AAG41465 standard; protein; 186 AA.
XX
XX  AAG41465;
XX
DT  18-OCT-2000 (first entry)
XX
DE  Arabidopsis thaliana protein fragment SEQ ID NO: 51593.
XX
KW  Protein identification; signal transduction pathway; metabolic pathway;
KW  hybridisation assay; Genetic mapping; gene expression control; promoter;
KW  termination sequence.
XX
OS  Arabidopsis thaliana.
XX
PN  EP033405-A2.
XX
PD  06-SEP-2000.
XX
PF  25-FEB-2000; 2000EP-00301439.
XX
PR  25-FEB-1999; 99US-0121825P.
PR  05-MAR-1999; 99US-0123180P.
PR  09-MAR-1999; 99US-0123548P.
PR  23-MAR-1999; 99US-0125788P.
PR  25-MAR-1999; 99US-0126264P.
PR  29-MAR-1999; 99US-0126785P.
PR  01-APR-1999; 99US-0127462P.
PR  06-APR-1999; 99US-0128224P.
PR  08-APR-1999; 99US-0128714P.
PR  16-APR-1999; 99US-0129845P.
PR  19-APR-1999; 99US-0130077P.
PR  21-APR-1999; 99US-0130449P.
PR  23-APR-1999; 99US-0130510P.
PR  28-APR-1999; 99US-0130891P.
PR  30-APR-1999; 99US-0131449P.
PR  30-APR-1999; 99US-0132048P.
PR  04-MAY-1999; 99US-0132407P.
PR  05-MAY-1999; 99US-0132484P.
PR  06-MAY-1999; 99US-0132485P.
PR  06-MAY-1999; 99US-0132486P.
PR  07-MAY-1999; 99US-0132487P.
PR  11-MAY-1999; 99US-0132863P.
PR  14-MAY-1999; 99US-0134256P.
PR  14-MAY-1999; 99US-0134218P.
PR  14-MAY-1999; 99US-0134219P.
PR  14-MAY-1999; 99US-0134219P.
PR  14-MAY-1999; 99US-0134221P.
PR  18-MAY-1999; 99US-0134370P.
PR  18-MAY-1999; 99US-0134768P.
PR  19-MAY-1999; 99US-0134941P.
PR  20-MAY-1999; 99US-0135124P.
PR  21-MAY-1999; 99US-0135353P.
PR  24-MAY-1999; 99US-0135629P.
PR  25-MAY-1999; 99US-0136021P.
PR  27-MAY-1999; 99US-0136392P.
PR  28-MAY-1999; 99US-0136782P.
PR  01-JUN-1999; 99US-0137422P.
PR  03-JUN-1999; 99US-0137528P.
PR  04-JUN-1999; 99US-0137502P.
PR  07-JUN-1999; 99US-0137724P.
PR  08-JUN-1999; 99US-0138094P.
PR  10-JUN-1999; 99US-0138540P.
PR  10-JUN-1999; 99US-0138847P.
PR  14-JUN-1999; 99US-0139119P.
PR  16-JUN-1999; 99US-0139452P.
PR  16-JUN-1999; 99US-0139453P.
PR  17-JUN-1999; 99US-0139492P.
PR  18-JUN-1999; 99US-0139454P.
PR  18-JUN-1999; 99US-0139455P.
PR  18-JUN-1999; 99US-0139456P.
PR  18-JUN-1999; 99US-0139457P.
PR  18-JUN-1999; 99US-0139458P.
PR  18-JUN-1999; 99US-0139459P.
PR  18-JUN-1999; 99US-0139460P.
PR  18-JUN-1999; 99US-0139461P.
PR  18-JUN-1999; 99US-0139462P.
PR  18-JUN-1999; 99US-0139463P.
PR  18-JUN-1999; 99US-0139750P.
PR  18-JUN-1999; 99US-0139763P.
PR  21-JUN-1999; 99US-0139817P.
PR  22-JUN-1999; 99US-0139899P.
PR  23-JUN-1999; 99US-0140353P.
PR  23-JUN-1999; 99US-0140354P.
PR  24-JUN-1999; 99US-0140695P.
PR  28-JUN-1999; 99US-0140823P.
PR  29-JUN-1999; 99US-0140991P.
PR  30-JUN-1999; 99US-0141187P.
PR  01-JUL-1999; 99US-0141842P.
PR  01-JUL-1999; 99US-0142154P.
PR  02-JUL-1999; 99US-0142055P.
PR  06-JUL-1999; 99US-0142390P.
PR  08-JUL-1999; 99US-0142803P.
PR  09-JUL-1999; 99US-0142920P.
PR  12-JUL-1999; 99US-0142977P.
PR  13-JUL-1999; 99US-0143542P.
PR  14-JUL-1999; 99US-0143624P.
PR  15-JUL-1999; 99US-0144005P.
PR  16-JUL-1999; 99US-0144085P.
PR  16-JUL-1999; 99US-0144086P.
PR  19-JUL-1999; 99US-0144325P.
PR  19-JUL-1999; 99US-0144331P.
PR  19-JUL-1999; 99US-0144332P.
PR  19-JUL-1999; 99US-0144333P.
PR  19-JUL-1999; 99US-0144334P.
PR  20-JUL-1999; 99US-0144352P.
PR  20-JUL-1999; 99US-0144632P.
PR  20-JUL-1999; 99US-0144884P.
PR  21-JUL-1999; 99US-0144814P.
PR  21-JUL-1999; 99US-0145086P.
PR  21-JUL-1999; 99US-0145088P.
PR  22-JUL-1999; 99US-0145085P.
PR  22-JUL-1999; 99US-0145087P.
PR  22-JUL-1999; 99US-0145089P.
PR  22-JUL-1999; 99US-0145102P.
PR  23-JUL-1999; 99US-0145145P.
PR  23-JUL-1999; 99US-0145218P.
PR  23-JUL-1999; 99US-0145224P.
PR  26-JUL-1999; 99US-0145276P.
PR  27-JUL-1999; 99US-0145913P.
PR  27-JUL-1999; 99US-0145918P.
PR  27-JUL-1999; 99US-0145919P.
PR  28-JUL-1999; 99US-0145951P.
PR  02-AUG-1999; 99US-0146386P.
PR  02-AUG-1999; 99US-0146388P.
PR  03-AUG-1999; 99US-0146389P.
PR  04-AUG-1999; 99US-0147038P.
PR  04-AUG-1999; 99US-0147204P.
PR  04-AUG-1999; 99US-0147302P.
PR  05-AUG-1999; 99US-0147192P.
PR  05-AUG-1999; 99US-0147260P.

```


CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 CC
 XX Sequence 524 AA;

Query Match 10.3%; Score 166; DB 6; Length 524;
 Best Local Similarity 23.1%; Pred. No. 0.00016;
 Matches 84; Conservative 59; Mismatches 143; Indels 78; Gaps 15;

QY 8 QDIQECTTQIQ-----YLKQVQPSVAQLRSTWDPAINLFFL-----KMKGELEQT 54
 DB 88 KDIALQERIEKREDTIQKQAREAQVNTSSNYIDAVLNADSLADAGRVOQMTWVKAN 147
 QY 55 KDKLEQAQNELSAWKFTPDSTQTKMLMAKRMILQENBELGRLSQRIQAEL-ALQ 113
 DB 148 NDLMEQKQDKKA---VEDKKAEND--AKLKELAEAOAALSEQ--KGDLLSKQADLNVLK 200
 QY 114 KKYSELKSSODELNDFTIQLDEEVEGMSQSTLVLLQQLKETRQQLAQ-----YQOQQ 166
 DB 201 TSLAEQNTAEKDKADLNQ-Q-KAEAEQARIREQORLAEGARQQAQAEKAEQAE 259
 QY 167 SQASAPSTRRTTASEPVEQSEATSKDCSRLTNGPNSGSSRSQRTSGSGFHREGNTEDDF 226
 DB 260 AEAQATQASSTAQSSATESSATQSSMTTESSATQSSATEST-----PESSTEESTA 314
 QY 227 PSSPFGNKNKSNSSEERTGRGSGYVNOJLSAGYESVDSPSGNSLTHOSN-----DT 279
 DB 315 PES--SATEESTTAPESATEESTTVPESATEESTTVP-----ESSTTEESTTAPPTPST 369
 QY 280 DSSHD-----POEEKAV-----SGKGN-----RTVGSRRHYONG 307
 DB 370 DQSVDTGNGTSGSTPAPTPTTPEQPKVTPAPAPSGSVNGAIAVAEAYKIGTPYVWGG 429
 QY 308 LDSS 311
 DB 430 KDPS 433

RESULT 13

ADCS95468
 ID ADCS95468 standard; protein; 525 AA.

XX
 AC ADCS95468;

XX 01-JAN-2004 (first entry)

DE B. faecium protein sequence SEQ ID 5095.

XX Vaccine; urinary tract infection; bacteraemia; endocarditis; wound;
 KW abdominal-pelvic infection.

XX Enterococcus faecium.

PN US583275-B1.

XX 24-JUN-2003.

XX 30-JUN-1998; 98US-00107532.

XX 02-JUL-1997; 97US-0051571P.

PR 14-MAY-1998; 98US-0085598P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Doucette-Stamm LA, Bush D;

XX WPI: 2003-799836/75.

XX N-PSDB; ADC91814.

XX New isolated nucleic acid derived from *Enterococcus faecium* encoding an
 PT *Enterococcus faecium* polypeptide useful for detection, prevention and
 PT treatment of a pathological condition resulting from a bacterial
 PT infection.

PS Example 1; SEQ ID NO 5095; 243pp; English.

XX The invention relates to an isolated nucleic acid derived from
 CC *Enterococcus faecium* encoding an *Enterococcus faecium* polypeptide having
 CC one of 10 fully defined sequences given in the (or comprising 40
 CC sequential nucleotides chosen from any of the nucleic acids, its
 CC complement or sequences hybridising to it). Also included are a
 CC recombinant vector comprising the nucleic acid operably linked to
 CC transcription regulatory element, a cell comprising the vector and a
 CC single-stranded probe comprising the nucleic acid. The nucleic acids are
 CC chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.
 CC The nucleic acids is useful for diagnosing pathological conditions
 CC resulting from *E. faecium* bacterial infection (e.g. urinary tract
 CC infection, bacteraemia, endocarditis, wounds and abdominal-pelvic
 CC infection) and for screening drugs such as agonists and antagonists. The
 CC nucleic acid is useful for recombinant production of *Candida albicans*.
 CC derived peptides or antisense polypeptides. Pharmaceutical compositions
 CC and vaccines containing the nucleic acid are useful for preventing or
 CC treating *Enterococcus faecium* infections. The present sequence represents
 CC one if the disclosed *E. faecium* proteins.

XX Sequence 525 AA;

Query Match 10.3%; Score 166; DB 7; Length 525;

Best Local Similarity 23.1%; Pred. No. 0.00016;

Matches 84; Conservative 59; Mismatches 143; Indels 78; Gaps 15;

QY 8 QDIQECTTQIQ-----YLKQVQPSVAQLRSTWDPAINLFFL-----KMKGELEQT 54

DB 89 KDIALQERIEKREDTIQKQAREAQVNTSSNYIDAVLNADSLADAGRVOQMTWVKAN 148

QY 55 KDKLEQAQNELSAWKFTPDSTQTKMLMAKRMILQENBELGRLSQRIQAEL-ALQ 113

DB 149 NDLMEQKQDKKA---VEDKKAEND--AKLKELAEAOAALSEQ--KGDLLSKQADLNVLK 201

QY 114 KKYSELKSSODELNDFTIQLDEEVEGMSQSTLVLLQQLKETRQQLAQ-----YQOQQ 166

DB 202 TSLAEQNTAEKDKADLNQ-Q-KAEAEQARIREQORLAEGARQQAQAEKAEQAE 260

QY 167 SQASAPSTRRTTASEPVEQSEATSKDCSRLTNGPNSGSSRSQRTSGSGFHREGNTEDDF 226

DB 261 AEAQATQASSTAQSSATESSATQSSMTTESSATQSSATEST-----PESSTEESTA 315

QY 227 PSSPFGNKNKSNSSEERTGRGSGYVNOJLSAGYESVDSPSGNSLTHOSN-----DT 279

DB 316 PES--SATEESTTAPESATEESTTVPESATEESTTVP-----ESSTTEESTTAPPTPST 370

QY 280 DSSHD-----POEEKAV-----SGKGN-----RTVGSRRHYONG 307

DB 371 DQSVDTGNGTSGSTPAPTPTTPEQPKVTPAPAPSGSVNGAIAVAEAYKIGTPYVWGG 430

QY 308 LDSS 311

[illegible]

Tue Nov 16 05:47:32 2004

SQ Sequence 897 AA;

Query Match	9.7%; Score 156; DB 2; Length 897;
Best Local Similarity	23.7%; Pred.No. 0.0021;
Matches	79; Conservative 56; Mismatches 101; Indels 98; Gaps 16;
Qy	3 LAIKGQDIQBCITCIQYLK-QVQOPSV- ---AQLRSTMDVPAINLFFLKMKGELEQFKDK 57
Db	356 LKXEDTVKQRTSEVQDQDEVRESINLKLQAKQOVQ- ---ELGELDEKQAK 407
Qy	58 LEQAQNELSAWKFTPDQCTGKKMAKCRMIIQENQELGRLSGRIIAQLFAELALQK- --- 114
Db	408 LEEQLQE- ---VRKKC- ---ABEAQL- ---ISSLKAEITTSQSQI 440
Qy	115 -KYSEELKSGDEIANDF- ---IIQDDEVGQVOSTILVLQQLKETRQLQACYQQQQSQAS 170
Db	441 SSYEEELKKAELSRLQQETAQLEESVSGKAQLFLOQHLOESQGEISMQMRLEM- - 498
Qy	171 APSRTITASEPVPQSEATSKDCSELTPGNSGSSSRQRTSGSFHREGNTTDDFFSSP 230
Db	499 - - - - -XDL-ETDNNQGNWSSPQSVLVNG- - - - -ATDYCSLS 529
Qy	231 GNGKSGNSSBERTCRGSGGYNQLSAGYESVDSTGSENSLTHQSNDDTSSHPDQBEKA 290
Db	530 TSSETANFNEHAEGQ- ---NNLES- ---BPTHQESSV-RSSPEIAPSDVTDESEA 576
Qy	291 VSGKGNRTVGR- ---HVQNGLDSSVNVQGSVL 319
Db	577 VTVAGNEKVTPRFDDDKHSKE- -BDPPFNVESSSL 608

Search completed: November 14, 2004, 12:33:14
Job time : 160 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2004, 12:33:24 ; Search time 40 Seconds
(without alignments)
767.329 Million cell updates/sec

Title: US-10-030-389-2

Perfect score: 1608

Sequence: 1 MRLATKGQDIQECTTQIYLV.....GSRHVQGLDSSVNVQGSVL 319

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 79.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	282	17.5	330	2 T47576	FKBP12 interacting
2	156.5	9.7	507	2 S05542	hypothetical prote
3	156	9.7	897	2 A54696	EGF receptor subst
4	152.5	9.5	1589	2 T13606	hypothetical prote
5	151	9.4	2288	2 T29999	hypothetical prote
6	150	9.3	456	2 E86903	hypothetical prote
7	150	9.3	2251	2 T24490	hypothetical prote
8	148	9.2	543	1 QFMSL	neurofilament trip
9	145	9.0	1938	1 A40997	myosin heavy chain
10	144.5	9.0	859	2 D96502	hypothetical prote
11	144	9.0	461	2 JN0097	secreted 45k prote
12	144	9.0	1807	2 S03124	vitellogenin A2 pr
13	141.5	8.8	697	2 T03834	nuclear distributi
14	140.5	8.7	461	2 H84099	cell wall-binding
15	140	8.7	1031	1 A38713	kinesin heavy chai
16	139.5	8.7	869	2 A88710	protein C43d2.2 li
17	139	8.6	1017	2 PC4035	cell-cycle-depende
18	138	8.6	355	2 C39725	hypothetical prote
19	138	8.6	577	1 A41289	moesin - human
20	138	8.6	2245	2 T18278	myosin heavy chain
21	137.5	8.6	782	2 A34218	Bic-D protein - fr
22	137	8.5	544	2 S07144	neurofilament trip
23	136	8.5	1837	2 T41023	probable nuclear p
24	135.5	8.4	472	2 S41720	intermediate filam
25	135.5	8.4	1509	1 A27224	myosin heavy chain
26	135.5	8.4	2139	2 T18296	myosin heavy chain
27	135	8.4	892	2 T50986	related to transcr
28	135	8.4	1436	2 S57238	forked protein 5.4
29	135	8.4	1449	2 S57237	forked protein 5.6

keratin 10, type I
hypothetical prote
early endosome ant
myosin heavy chain
keratin, 67k type
hypothetical prote
myosin-like protei
vitellogenin 1 - m
epidermal growth f
probable transcrip
myosin heavy chain
keratin, 54k type
hypothetical prote
smooth muscle myos

30 134.5 8.4 561 2 A31994
31 134.5 8.4 782 2 T32155
32 134.5 8.4 1410 1 A57013
33 134.5 8.4 1940 2 A59287
34 134 8.3 645 2 A44861
35 133 8.3 613 2 S48557
36 132.5 8.2 1054 2 D96519
37 132 8.2 790 2 T05576
38 132 8.2 930 2 T08588
39 132 8.2 1704 2 T43141
40 131.5 8.2 896 2 S43074
41 131 8.1 662 2 T18233
42 130.5 8.1 484 2 B33501
43 130.5 8.1 526 1 KEB0VI
44 130.5 8.1 967 2 S66852
45 130.5 8.1 1938 2 JC5421

ALIGNMENTS

RESULT 1

T47576
FKBP12 interacting protein (FIP37) - Arabidopsis thaliana
N:Alternate names: Protein F24B22.130
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 09-Jul-2004
C:Accession: T47576
R:Blöcker, H.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.; Quettier, F.; Salanoubat, M.
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z23016
A:Accession: T47576
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-330 <BL0>
A:Cross-references: UNIPROT:Q9ZS28; EMBL:AL132957
A:Experimental source: cultivar Columbia; BAC clone F24B22
C:Genetics:
A:Map position: 3
A:Introns: 24/1; 60/2; 72/3; 102/1; 123/3; 149/3; 165/3; 183/3; 237/3; 261/1; 279/3
A:Note: F24B22.130

Query Match 17.5%; Score 282; DB 2; Length 330;

Best Local Similarity 33.5%; Pred. No. 5e-09;

Matches 55; Conservative 51; Mismatches 58; Indels 0; Gaps 0;

QY 1 MRLATKGQDIQECTTQIYLVQVQPSVAQLRSTWDPAINLFLKMKGELEQTKDKLEQ 60

Db 136 VOYAKREQEMAEIKSAVRDLKSQLKPSNQARLLDPAIHFEFSLKLVVEEKDKKIKE 195

QY 61 AQNELSAWKFTPDSTGKMLAKRMILQENQELGRLSQGRIAQLEALQKKYSBEL 120

Db 196 LQDNLAAVTFQSGKNGKMLAKCTLOEENEIGHQABGKIHELAKLAWKQSQNAEL 255

QY 121 KSSQDELNDFFIQLDEEVGMOSTILVLOQQLKETRQQLAQYQQ 164

Db 256 RSQFEGLYKHMEELNDVRSNETVILQELKEKEKSIERVKK 299

RESULT 2

S05542

hypothetical protein, 54K - Enterococcus faecium

C:Species: Enterococcus faecium

C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 09-Jul-2004

C:Accession: S05542

R:Fuerst, P.; Moesch, H.U.; Solioz, M.

Nucleic Acids Res. 17, 6724, 1989

A:Title: A protein of unusual composition from Enterococcus faecium.

A:Reference number: S05542; MUID:89385998; PMID:2780297

A:Accession: S05542

A:Molecule type: DNA

A:Residues: 1-507 <FUE>

A:Cross-references: UNIPROT:PI3692; GB:X16421; EMBL:M26048; NID:G43333; PIDN:CAA34442.1

```

Query Match          9.7%; Score 156.5; DB 2; Length 507;
Best Local Similarity 22.4%; Pred. No. 0.087;
Matches 84; Conservative 50; Mismatches 118; Indels 123; Gaps 14;

QY      11 QECTTQIYLKQVQSPVAQLRSTWMDPAINLFFL----- 45
Db      91 REATIO-----KQARETQVKNTSNYIDAVLNADSLADAVGRIQAMSTIVKANODLVQOQK 146

QY      46 KMKGELEQTKDKLEQKQNELSANKFTPDSTQTKKLMAKCRMLIQENQLGRQLSQGRIAQ 105
Db      147 EDQAQVEAKAEAKAEKQELADNQAALESQKG-DLLAK-----QADLVN 189

QY      106 LEAEALQKRYSELKSSODELNDFTIQLDEEVEGMSSTILVLQOQLKETRQQLAQVQ-- 163
Db      190 LKTSLAEEQATAEDKKA--DLN---RKKAEEAEQARIREQARLAEQARQAQAEKAE 242

QY      164 ---QQQSQASAPST-----SRTTASEPVEOS-----EATSKCGRSLTNGP 200
Db      243 KEAREQAQAQAQOALSSASTTIESSAQSSSESKAPESSTTBESSTEGSTTENS 302

QY      201 SNGSSSRQRTSGSGFHREGNTTEDDPP-----SSPCNGKNKSSNSEERTGRGSGGYVWQ 254
Db      303 STGSSSTESS-----TEBSTVPESSTQBSTPANTESSSSSSNTVNNNTNNTN- 352

QY      255 LSAGYESVDSPTCSENLSLTHQSDNTDS-----SHDPOEKAVSG-----KGN 296
Db      353 -----NSTNNSTNNNNNNNTVTAPTPTTPAPAPNPSPGSGVNGAAIVAEAY 401

QY      297 RTVGSRRHVQGLDSS 311
Db      402 KYVIGTPYVWGKDPS 416

RESULT 3
AS4696
EGF receptor substrate eps15 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change 09-Jul-2004
C:Accession: AS4696
R:Razjoli, F.; Minichiello, L.; Matoskova, B.; Wong, W.T.; Di Fiore, P.P.
Mol. Cell. Biol. 13, 5814-5829, 1993
A>Title: eps15, a novel tyrosine kinase substrate, exhibits transforming activity
A:Reference number: AS4696; MUID:93361014; PMID:7689153
A:Accession: AS4696
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-897 <FAZ>
C:Keywords: EF hand; phosphoprotein
F:48-80/Domain: calmodulin repeat homology <EF1>
F:160-192/Domain: calmodulin repeat homology <EF2>
F:723-253/Domain: calmodulin repeat homology <EF3>

Query Match          9.7%; Score 156; DB 2; Length 897;
Best Local Similarity 23.7%; Pred. No. 0.18;
Matches 79; Conservative 56; Mismatches 101; Indels 98; Gaps 16;

QY      3  LATKQDQIQECTTQIYLK-QVQQPSV-----AQLRSTWMDPAINLFFLKMKELEQTKDK 57
Db      356 LKEEDTQKRTSEVQDQDEVRESINLQKLQAKQVQ-----ELLGELDEQKAAQ 407

QY      58 LEQAQNELSAMKFTPDSTQTKKLMAKCRMLIQENQLGRQLSQGRIAQLEAPLALQK-- 114
Db      408 LEEQLQE-----VRKKC--AEEAQL-----ISSLKAEITSQESQI 440

QY      115 -KYSSELKSSODELNDF---ITQLDEEVEGMSSTILVLQOQLKETRQQLAQVQOQSSAS 170
Db      441 SSYEELLKAEELSRLOETRAQLAESVSGKAQLEPQQHLQESQQEITSSNMPLFM-- 498

QY      171 APSTSRTTASFPVEQSEATSKDCSLRTNGPNSSSSRQTSQSGPHREGNTTEDDPPSP 230

```

```
RESULT 5
T29999
Hypothetical protein ZC8.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T29999
R:Atreille, P.; Bradshaw, H.
submitted to the EMBL Data Library, July 1996
A:Description: The sequence of C. elegans cosmid ZC8.
A:Reference number: Z20719
A:Accession: T29999
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2288 <LAT>
A:Cross-references: UNIPROT:Q23081; EMBL:U64862; PIDN:AA52624.1; GSPDB:GN00028; CESP:T29999
A:Experimental source: strain Bristol N2; clone ZC8
C:Genetics:
A:Gene: CESP:ZC8.4
A:Map position: X
A:Introns: 43/2; 80/3; 110/3; 153/2; 274/3; 738/2; 1249/1; 1392/3; 1454/1; 1602/1; 1696/1
Query Match 9.4%; Score 151; DB 2; Length 2288;
Best Local Similarity 21.6%; Pred. No. 0.99;
Matches 79; Conservative 54; Mismatches 137; Indels 96; Gaps 13;
QY 10 IQECTTQIYLVKQV---QPSVAQLRSTWVDPAINLFFLMKGELEQTQKLEQAQNELS 66
DB 671 IEEKETQIRYSDDIRNICKDLDLREKY--DRVHTDNEKILGELSHAQKAHAHAEQQKL 728
QY 67 AKWFTDPS-----QTKKGLWA 82
DB 729 EIKIQRDDYQKQDEHARHFLDIRHKLETEIKGRQLEKNGKARNDELKLRQTTSIDYES 788
QY 83 KCRMLIOENQELGRQLS--QGRIQAQAEALQKQYSEEL-----KSSQDELNDFTI 132
DB 789 QINLLRHNDELDTTIKGHQKITHLENELHSRSGEIKLNDLNQRLQKEKQDILNQ-KL 847
QY 133 OLDEEVGQSTILVLOOQLKEFRQLAQY-----QQQSCASAPSTSRRTASE 181
DB 848 KLPGDVQALKEITKLENELEKLNENKELVGEARDAANQOLGRANLLNKELEDTKQ 907
QY 182 PVEQSEATSKDC-----SRLTNGPSNGSSSRQRTSGSGFRGNTTDDFPSPFGNG 233
DB 908 DLKHSTDVNQLEQIDRLKXERLANIKGGRISRDSITGI-----DGGAP-----G 953
QY 234 NKSSNSEERTGRGGGYVNLQAGYVSDSPGSGNSLTHQNDTDSHDPQEEKAVSG 293
DB 954 DRSSVADPSRT-RGAAGSTVFVPA--EDIESRGGGEIDIP--SSGDVIHGRDGRDAGN 1010
QY 294 KGNRTV 299
DB 1011 RGHWTI 1016
RESULT 6
E86903
Hypothetical protein usp45 [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 09-Jul-2004
C:Accession: E86903
R:Botolin, A.; Winkler, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s8
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: E86903
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-456 <STO>
A:Cross-references: UNIPROT:Q9CJD1; GB:AB005176; PID:g12725236; PIDN:AAK06327.1; GSPDB:G
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: usp45
```

```
Query Match 9.3%; Score 150; DB 2; Length 456;
Best Local Similarity 24.9%; Pred. No. 0.18;
Matches 73; Conservative 47; Mismatches 103; Indels 70; Gaps 13;
QY 2 ELATKGQDIOECTTQIYLVKQVQPSVAQLRSTWVDPAINLFFL-----KMKGE 50
DB 88 QIATLNESIAERTKTLB--AQRSAQVNSATNMDAVNSKSLTDVIQKTAIATVSSA 145
QY 51 LEQTKDLKLEQAQNELSAWKFTPDQCTGK---LMAKCRMLIOENQELGRQLSGRIACL 106
DB 146 NKQMLEQKEQEKELS-----QKSETVKKNYNQVSLSQSLDSQAQELTQQAEELKVATL 200
QY 107 EAEIALQKQYSEELKSSQDELNDFTIOLDEEVGQSTILVLOOQLKETQQLAQYQQQQ 166
DB 201 -----NQQTATATAQDKKQSL---LDEKA-----AAEKAQAQAAKKQAYEAQ 241
QY 167 SQAS---APSTSRIT-----ASEPVEQSEATSKDCSRLTNGPSNGSSSRQRTSGSGPHRE 218
DB 242 KEAKAQAQAAATAATTKAVEEATSVSSQA--SQSSSSSSNTSSNTSSSSSSSN--- 296
QY 219 GNTTDDFPSPFGNGKSSNSEERTGRGGSGYVNLQAGYVSDSPSGSENS 271
DB 297 -----SSSSSSSSSSSS---NGGGSTNTGNNNAAG---TGNTGGSSS 333
RESULT 7
T24490
Hypothetical protein T05A10.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T24490
R:Sulston, J.
submitted to the EMBL Data Library, November 1995
A:Reference number: Z19898
A:Accession: T24490
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2251 <WTL>
A:Cross-references: UNIPROT:Q22190; EMBL:Z68108; PIDN:CAA92133.1; GSPDB:GN00028; CESP:T.
A:Experimental source: clone T05A10
C:Genetics:
A:Gene: CESP:T05A10.1
A:Map position: X
A:Introns: 188/3; 240/3; 420/1; 570/3; 596/1; 732/3; 778/3; 851/3; 1359/2; 1394/2; 1434
Query Match 9.3%; Score 150; DB 2; Length 2251;
Best Local Similarity 20.5%; Pred. No. 1.1;
Matches 82; Conservative 68; Mismatches 144; Indels 106; Gaps 15;
QY 4 ATKGDQIQ-----ECTTQIYLVKQVQPSVAQLRSTWVDPAINL-----FFLKM 47
DB 248 AQAQAHVQSRQVQPSQSQVQAQLQQQQLQQQAQQLSQQAQQQQLQQLQQLQQ 307
QY 48 KGELEQTQKLEQAQNELSAWKFTPD-----SOTGKKLMKCRMLIOEN 91
DB 308 QQQLHQQRAAQAQQAQNAQNSAQQRPSVASTPALSTPOLNDLTQTMQALQQ--QLLQQ 366
QY 92 QBLGRQLSGRIQAQAEIALQKQYSEELKSSQDELNDFTI--IQLDEEVG---MQSTIL 146
DB 367 QAQAQAQAQAQAQL--AQAQQAQQQSQNRVTSQALQYIQSMQLQQRADGTENAESQEE 425
QY 147 VLOQQLKETQQLAQYQQQQSQ----- 168
DB 426 RLQAQMLNEQQQRLQQAQAEAHQRLQLLSSTPAPRGGITMGTPTIGIARREQPVSTVAV 485
QY 169 ASAPSTSRIT--TASEPVEQSEATSKDCSRLTNGPSNGSSSRQRTSGSGFRGNTTDDFP 227
DB 486 TTAAPAVRTPVAVPMKQNSPNM-----PSSTSTISASATSSHQILAPLSKPLQEP 538
QY 228 SS---PGNGKSSNS-----SEERTGRGGSGYVNLQAGYVSDSPGSGNSLTHQSN- 277
DB 539 SSSKAASSGNSGMSDHISRIISENEVILQGDPIVRKGRPYHRQI---GAQSSVDHDSNS 594
```

QY 278 --DTSDSHDPOEKAVSGKN-----RTVGRHVQNGLDS 310
 Db 595 GGSRTSPGPKDSRMLQAAASRSQSLFELSGSKHFMGLSLS 634

RESULT 8
 QPMSL
 neurofilament triplet L protein - mouse
 N:Alternate names: 68K neurofilament protein; NF-L(low) protein; type IV IF protein
 C:Species: Mus musculus (house mouse)
 C>Date: 31-Mar-1998 #sequence revision 30-Sep-1993 #text_change 09-Jul-2004
 C:Accession: A25227; A26562; A43772; A41012; I55316
 R:Lewis, S.A.; Cowan, N.J.
 Mol. Cell. Biol. 6, 1529-1534, 1986
 A>Title: Anomalous placement of introns in a member of the intermediate filament multigene family
 A:Reference number: A25227; MUID:87064433; PMID:3785173
 A:Accession: A25227
 A:Molecule type: DNA
 A:Residues: 1-543 <LEW>
 A:Cross-references: UNIPROT:P08551; GB:M13016; NID:G200023; PIDN:AAA39810.1; PID:G387492
 A>Note: the authors translated the codon GGC for residue 5 as Ala, ACA for residue 88 as 1 as Glu
 R:Lewis, S.A.; Cowan, N.J.
 J. Cell Biol. 100, 843-850, 1985
 A>Title: Genetics, evolution, and expression of the 68,000-mol-wt neurofilament protein:
 A:Reference number: A26562; MUID:8511334; PMID:3919033
 A:Accession: A26562
 A:Molecule type: mRNA
 A:Residues: 242-543 <LE2>
 A:Cross-references: GB:X02165
 A:Experimental source: brain
 R:Julien, J.P.; Meyer, D.; Flavell, D.; Hurst, J.; Grosfeld, F.
 Brain Res. Mol. Brain Res. 1, 243-250, 1986
 A>Title: Cloning and developmental expression of the murine neurofilament gene family.
 A:Reference number: A43772
 A:Accession: A43772
 A:Molecule type: mRNA
 A:Residues: 1-5, 'Y', '7-8, 'Y', '10-64, 'M', '66-72, 'L', '74-98, 'D', '100-194, 'R', '196-202, 204-239, 'Y'
 A:Cross-references: GB:M20480; NID:G200037; PIDN:AAA39814.1; PID:G200038
 A>Note: the authors translated the codon CGC for residue 195 as Ala
 R:Shag, R.K.; Nixon, R.A.
 J. Biol. Chem. 266, 18861-18867, 1991
 A>Title: Identification of Ser-55 as a major protein kinase A phosphorylation site on the
 A:Reference number: A41012; MUID:92011653; PMID:1717455
 A:Accession: A41012
 A:Molecule type: Protein
 A:Residues: 52-57 <STH>
 R:Nakahira, K.; Ikenaka, K.; Wada, K.; Tamura, T.
 J. Biol. Chem. 265, 19786-19791, 1990
 A>Title: Structure of the 68-kDa neurofilament gene and regulation of its expression.
 A:Reference number: I55316; MUID:91060592; PMID:2246261
 A:Accession: I55316
 A>Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-5, 'Y', '7-8, 'Y', '10-28 <RES>
 A:Cross-references: GB:M55423; NID:G200027; PIDN:AAA39812.1; PID:G554245
 C:Comment: This is the most abundant of the three neurofilament proteins and, as the oth
 C:Genetics:
 A:Introns: 349/3; 391/2; 498/1
 C:Superfamily: cytoskeletal keratin
 C:Keywords: coiled coil; intermediate filament
 F:2-72/Domain: head <HED>
 F:94-125/Domain: coil 1a, alpha-helical rod #status predicted <R1A>
 F:126-138/Domain: linker 1
 F:139-234/Domain: coil 1b, alpha-helical rod #status predicted <R1B>
 F:235-256/Region: linker 12
 F:257-272/Domain: coil 2a, alpha-helical rod #status predicted <R2A>
 F:273-281/Region: linker 2
 F:282-401/Domain: coil 2b, alpha-helical rod #status predicted <R2B>
 F:404-543/Domain: tail <TAI>
 F:404-444/Region: tail subdomain a
 F:445-543/Region: tail subdomain b

Query Match 9.2%; Score 148; DB 1; Length 543;
 Best Local Similarity 19.3%; Pred. No. 0.28;
 Matches 69; Conservative 55; Mismatches 164; Indels 70; Gaps 9;

QY 4 ATKGDQ-----IQECTTQIQYLKQVQPSVAQLR-----STMVD-----PA 39
 Db 195 ARKGADAAALAAAELEKIDSIMDEIAFLKKVHEEIAELQAIQIAQISYEMDVSKPD 254

QY 40 INFLFKMKGLEOTPKLEQAQNELSAWKTPDSQTKLMKRCMLIQNQELGRLS 99
 Db 255 LSAALKDIPRAQYELAAKQWQAEWFKSRFTVLITESAKNTDAVRAKDEVSE-SRLL 313

QY 100 QGRIQAQLEALQKYSSEELKSSQDELNDFIOLDEEVEGQSTILVLOQLKETRQQL 159
 Db 314 KAKTLEIEACRGMNEALEKQLELEDKQN-----ADISAMQDTINKLENELESTKSEM 366

QY 160 AQY---QQQSCASAPSTRRTTASEPVEOSEATS---KDCSRLTNGPSNGSSSRQRTSGS 213
 Db 367 ARYLKEYQDLINVMALDIEIAAYRKLLGEETRLSTVSGSIISGYSQSSQVFEGRGAYS 426

QY 214 GFHR-----EGNTEDDFPSSPGNGKNSSEERT 244
 Db 427 GLQSSSYLMSARSPPAYTSHVQEQTEVEETIEATKAEAKDEPPSGEAEKEKEKE 486

QY 245 GRGGGYVQLSAGYVESDPTGSENLSITQSNDDTSDSHDPQBEKAYSGKGRVGSF 302
 Db 487 GEEEGAEBAKDESEDTKEEBEGG---EGEEEDTKSEEEKEEAGEEAVAKK 541

RESULT 9
 A40997
 myosin heavy chain, striated adductor muscle - scallop (Aequipecten irradians)
 N:Contains: myosin A1Pase (SC 3.6.4.1)
 C:Species: Aequipecten irradians
 C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
 C:Accession: A40997; S13557
 R:Nyitra, L.; Goodwin, E.B.; Szent-Gyorgyi, A.G.
 J. Biol. Chem. 266, 18469-18476, 1991
 A>Title: Complete primary structure of a scallop striated muscle myosin heavy chain. Seq
 A:Reference number: A40997; MUID:92011595; PMID:1917970
 A:Accession: A40997
 A:Molecule type: mRNA
 A:Residues: 1-1938 <NYI>
 A:Cross-references: UNIPROT:P24733; GB:X55714; NID:G5611; PIDN:CAA39247.1; PID:G5612
 C:Superfamily: myosin heavy chain; myosin motor domain homology
 C:Keywords: actin binding; ATP; coiled coil; hydrolase; muscle contraction; nucleotide b
 F:86-763/Domain: myosin motor domain homology <MMOT>
 F:176-183/Region: nucleotide-binding motif A (P-loop)
 F:547-586/Region: actin binding #status predicted
 F:653-675/Region: actin binding #status predicted
 F:836-1938/Domain: coiled coil #status predicted <COI>
 F:836-1276/Region: S2
 F:1277-1938/Region: light meromyosin
 F:182/Binding site: ATP (Lys) #status predicted
 F:693/703/Active site: Cys #status predicted

Query Match 9.0%; Score 145; DB 1; Length 1938;
 Best Local Similarity 21.1%; Pred. No. 1.8;
 Matches 72; Conservative 63; Mismatches 146; Indels 60; Gaps 11;

QY 3 LATKGDP-----IQECTTQIQYLKQVQPSVAQLRSTWDPAINLFF-LKKKGLEQTKDK 57
 Db 1189 LRKHDAANEMADQVDQLQKVKSKLEKDKKDLKRENDLESQMTNHNKNGSEVMKQ 1248

QY 58 LEQAQNELSAWKTPDSQTKLMKRCMLIQNQELGRL-----SQGRIQAQLEALQKK 115
 Db 1249 FESQMSPLNA-LEDSQRSINELQSKSRLOANSDLTQLQLEDAEHRVSVLSKEKQLSS 1307

QY 116 YSELKSSQDELNDFIOLDEEVEGQSTILVLOQLKETRQQLAQYQQQSOAS----- 170
 Db 1308 QLEDARRSLEETRAKSLQNEVRNHDMDAIREQEBEQESKSDVQRLSKANNEIQ 1367

QY 171 -----APSTSR-----TASBPVEQSEATSKDCSRLTNGPS----- 201

Tue Nov 16 05:47:33 2004

RESULT 15

A38713
 kinesin heavy chain - sea urchin (Strongylocentrotus purpuratus)
 N;Contains: Kinesin ATPase (EC 3.6.1.-)
 C;Species: Strongylocentrotus purpuratus (purple urchin)
 C;Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 09-Jul-2004
 C;Accession: A38713; S24308
 R;Wright, B.D.; Henson, J.H.; Wedaman, K.P.; Willy, P.J.; Morand, J.N.; Scholey, J.M.
 J. Cell Biol. 113, 817-833, 1991
 A;Title: Subcellular localization and sequence of sea urchin kinesin heavy chain: evidence
 A;Reference number: A38713; MUID:91225077; PMID:1827446
 A;Accession: A38713
 A;Molecule type: mRNA
 A;Residues: 1-1031 <WRI>
 A;Cross-references: UNIPROT:P35978; GB:X56844; NID:G10269; PIDN:CAA40175.1; PID:G10270
 C;Superfamily: kinesin heavy chain; kinesin motor domain homology
 C;Keywords: ATP; coiled coil; heterotetramer; hydrolase; microtubule binding; mitosis; n
 F;9-331/Domain: kinesin motor domain homology <KMT>
 F;84-91/Region: nucleotide-binding motif A (P-loop)
 F;393-857/Domain: rod #status predicted <RDD>
 F;858-1031/Domain: globular #status predicted <GBD>
 F;90/Binding site: ATP (lys) #status predicted

Query Match 8.7%; Score 140; DB 1; Length 1031;
 Best Local Similarity 21.3%; Pred. No. 1.6;
 Matches 70; Conservative 50; Mismatches 125; Indels 84; Gaps 11;

Qy	2	RLATKGDDIOECTTQIYLKQVQPSVAQLRSTWDPAINLFLKMK-----GEL 51
Db	710	QLANLRTEINEKEHQMBELXDNQ-----RMTLQHEKLQLDYEKLEAEAKAKREL 763
Qy	52	EOTKDKLEQACNELSANKTTPDSQTKKL--MAKCRMLI-----QENQE 93
Db	764	SOQDRREQAKQDLKLEET-----VAKELQTLNLRKLFVSDLQNRVKKALEGGDRDDDS 819
Qy	94	IGRLSQGRIFAQLEALQKYSSEELKSSQDEINDFIIQLDEVEGMQSTILVLOQQLK 153
Db	820	GGSAQAKQKISFLNNLEQLTKVHKQLVRDNADLRCELKLERLRATSERVKALEMSLK 879
Qy	154	ETRO-----QLAQYQQQSQAAPSTSTTASEPVEQSEATSKDCSRLTNGSPNGSSSRQR 209
Db	880	ETKEGAMRDRKRYQOEVDRI-----REAVRORNFAPKRGSSAQIA 918
Qy	210	TSGSGFHREGNTTDDFPSPGNGKSNSEERTGRGGSGYVQLSAGYVESVDSPT-GS 268
Db	919	KAIRAGHP-----PPSPGGS-----TGIRGGGSGIRGGGSPVIRPPPSHGS 959
Qy	269	ENSLTHOSNDTSSHDPOEEKAVSGKGNR 297
Db	960	PEPISH-NNSPEKSLNPDAENMEKANK 987

Search completed: November 14, 2004, 12:46:19
 Job time : 43 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2004, 12:36:39 ; Search time 141 Seconds
(without alignments)
800.484 Million cell updates/sec

Title: US-10-030-389-2

Perfect score: 1608

Sequence: 1 MRLATKGDIQECTTQIQYL.....GSRHVQNGLDSSVNVQGSVL 319

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1568699 seqs, 353819137 residues

Total number of hits satisfying chosen parameters: 1568699

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

1: /cgn2_6/prodata/1/pubaa/US07_PUBCOMB.pep.*
2: /cgn2_6/prodata/1/pubaa/US06_PUBCOMB.pep.*
3: /cgn2_6/prodata/1/pubaa/US05_PUBCOMB.pep.*
4: /cgn2_6/prodata/1/pubaa/US04_PUBCOMB.pep.*
5: /cgn2_6/prodata/1/pubaa/US03_PUBCOMB.pep.*
6: /cgn2_6/prodata/1/pubaa/US02_PUBCOMB.pep.*
7: /cgn2_6/prodata/1/pubaa/US01_PUBCOMB.pep.*
8: /cgn2_6/prodata/1/pubaa/US00_PUBCOMB.pep.*
9: /cgn2_6/prodata/1/pubaa/US09_PUBCOMB.pep.*
10: /cgn2_6/prodata/1/pubaa/US08_PUBCOMB.pep.*
11: /cgn2_6/prodata/1/pubaa/US07_PUBCOMB.pep.*
12: /cgn2_6/prodata/1/pubaa/US06_PUBCOMB.pep.*
13: /cgn2_6/prodata/1/pubaa/US05_PUBCOMB.pep.*
14: /cgn2_6/prodata/1/pubaa/US04_PUBCOMB.pep.*
15: /cgn2_6/prodata/1/pubaa/US03_PUBCOMB.pep.*
16: /cgn2_6/prodata/1/pubaa/US02_PUBCOMB.pep.*
17: /cgn2_6/prodata/1/pubaa/US01_PUBCOMB.pep.*
18: /cgn2_6/prodata/1/pubaa/US00_PUBCOMB.pep.*
19: /cgn2_6/prodata/1/pubaa/US09_PUBCOMB.pep.*
20: /cgn2_6/prodata/1/pubaa/US08_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	357	22.2	151	9	US-09-840-787-1
2	351	21.8	188	17	US-10-425-115-191548
3	272	16.9	209	16	US-10-767-701-39876
4	269.5	16.8	376	17	US-10-425-115-307181
5	285.5	16.5	265	15	US-10-425-114-57475
6	264	16.4	343	15	US-10-424-599-240261
7	262.5	16.3	195	17	US-10-425-115-203631
8	256	15.9	213	15	US-10-425-114-48881
9	256	15.9	336	16	US-10-437-963-196776
10	256	15.9	352	14	US-10-310-154-593
11	247	15.4	190	15	US-10-425-114-55340
12	238	14.8	239	17	US-10-739-930-9291
13	238	14.8	310	15	US-10-424-599-240263

Sequence 240259, A
Sequence 42229, A
Sequence 208116, A
Sequence 57658, A
Sequence 468, App
Sequence 10932, A
Sequence 677, App
Sequence 30, Appl
Sequence 33, Appl
Sequence 29, Appl
Sequence 20, Appl
Sequence 305, App
Sequence 57375, A
Sequence 176140, A
Sequence 31, Appl
Sequence 2795, App
Sequence 4, Appl
Sequence 4, Appl
Sequence 4, Appl
Sequence 24, Appl
Sequence 113, App
Sequence 158, App
Sequence 57, Appl
Sequence 54, Appl
Sequence 27, Appl
Sequence 453, App
Sequence 454, App
Sequence 187, App
Sequence 25, Appl
Sequence 126, App
Sequence 89, Appl
Sequence 123248, A

14 234 14.6 300 15 US-10-424-599-240259
15 218.5 13.6 165 15 US-10-425-114-42229
16 203 12.6 146 15 US-10-424-599-208116
17 166 10.3 524 15 US-10-282-122A-57658
18 156.5 9.7 511 16 US-10-451-467A-468
19 152.5 9.5 1031 9 US-09-815-242-10932
20 151 9.4 2288 14 US-10-369-493-6774
21 150 9.3 431 16 US-10-383-930-30
22 149 9.3 431 16 US-10-383-930-33
23 146 9.1 431 16 US-10-383-930-29
24 144 9.0 716 14 US-10-114-774-20
25 142.5 8.9 689 13 US-10-108-605-305
26 142.5 8.9 1061 15 US-10-282-122A-57375
27 142.5 8.9 3161 16 US-10-437-963-176140
28 141.5 8.8 432 16 US-10-383-930-31
29 141 8.8 294 15 US-10-264-049-2795
30 141 8.8 1742 15 US-10-615-383-4
31 141 8.8 1742 16 US-10-690-184-4
32 141 8.8 1742 16 US-10-689-082-4
33 140 8.7 1031 14 US-10-080-608A-24
34 140 8.7 1031 14 US-10-370-685-113
35 139 8.6 1017 16 US-10-408-765A-158
36 138.5 8.6 542 15 US-10-205-331-57
37 138 8.6 577 14 US-10-236-031B-54
38 138 8.6 577 16 US-10-360-849A-27
39 138 8.6 577 16 US-10-408-765A-453
40 138 8.6 577 16 US-10-408-765A-454
41 138 8.6 577 16 US-10-408-765A-454
42 138 8.6 813 9 US-09-964-899-25
43 138 8.6 1165 14 US-10-021-660-126
44 138 8.6 1165 15 US-10-211-462-89
45 137 8.5 868 16 US-10-437-963-123248

ALIGNMENTS

RESULT 1

US-09-840-787-1
; Sequence 1, Application US/09840787
; Patent No. US20020058264A1
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; Hillman, Jennifer L.
; Bandman, Olga
; Shah, Purvi
; Au-Young, Janice
; Yue, Henry
; Guegler, Karl J.
; Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/840,787
; FILING DATE: 23-Apr-2001
; CLASSIFICATION: <Unknown>
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 09/518,865
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.

US-10-767-701-39876

Db 233 NVAANFTSSKLGKMLAKCRTLQENEEIGAMASEGKIHELGMKIAVLKQNNELRNQ 292
QY 124 QDELNDFFIQLDEEVEGMSSTILVLQOQLKETROQLAQYQOQOQ-SQASAPSTSRRTASEP 182
Db 293 FDVLYKHMGGVNDVERSNELVAILQEELEAKELEVLRLKEETLSCKGCTPEAPAPAP 352
QY 183 VEQSEATSKD 192
Db 353 VEEGDGAGND 352

RESULT 5

US-10-425-114-57475
; Sequence 57475, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 57475
; LENGTH: 285
; TYPE: PR1
; ORGANISM: Zea mays
; OTHER INFORMATION: Clone ID: UC-ZMFLB73123A07_FLI.pep
US-10-425-114-57475

Query Match 16.5%; Score 265.5; DB 15; Length 265;
Best Local Similarity 31.7%; Pred. No. 3.5e-10;
Matches 65; Conservative 47; Mismatches 80; Indels 13; Gaps 3;

QY 4 ATKQDIQECTTOIYLYKQVQPSVAQLRSTWVDPAINLFLKMKGELETKDLQQAQN 63
Db 69 AKREQEIAELKSAVRDLTKQLRPPSMQTRLLIDPAIHFEFTRLKNLVEEKKIKELQD 128
QY 64 ELSAWKTPDSOTGKIKMAKCRMLIQENELGRLSQGRIAGLEAEALQKYSEELKSS 123
Db 129 NVAANFTSSKLGKMLAKCRTLQENEEIGTMASEGKIHELGMKIAVLKQNNELRNQ 188
QY 124 QDELNDFFIQLDEEVEGMSSTILVLQOQLKETROQLAQYQOQOQ-SQASAPSTSRRTASEP 183
Db 189 FDVLYKHMGGVNDVERSNELVAILQEELEAKELEVLRLKEETLSCKGCTPEAPAP 242
QY 184 EQSEATSKDCSRLTNGPSNGSSSRQ 208
Db 243 E-----TSDDAG---NGQEADSDTLQ 260

RESULT 6

US-10-424-599-240261
; Sequence 240261, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684

; SEQ ID NO 240261
; LENGTH: 343
; TYPE: PR1
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_58983C.1.pep
US-10-424-599-240261

Query Match 16.4%; Score 264; DB 15; Length 343;
Best Local Similarity 32.7%; Pred. No. 6e-10;
Matches 54; Conservative 47; Mismatches 64; Indels 0; Gaps 0;

QY 4 ATKQDIQECTTOIYLYKQVQPSVAQLRSTWVDPAINLFLKMKGELETKDLQQAQN 63
Db 146 AKREQEIAELKSAVRDLTKQLRPPSMQTRLLIDPAIHFEFTRLKNLVEEKKIKELQD 205
QY 64 ELSAWKTPDSOTGKIKMAKCRMLIQENELGRLSQGRIAGLEAEALQKYSEELKSS 123
Db 206 NVAANFTSSKLGKMLAKCRTLQENEEIGTMASEGKIHELGMKIAVLKQNNELRNQ 265
QY 124 QDELNDFFIQLDEEVEGMSSTILVLQOQLKETROQLAQYQOQOQ-SQ 168
Db 266 FEGLQKHMGGVNDVERSNELVAILQEELEAKELEVLRLKEETLSCKGCTPEAPAP 310

RESULT 7

US-10-425-115-203631
; Sequence 203631, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 203631
; LENGTH: 195
; TYPE: PR1
; ORGANISM: Zea mays
; OTHER INFORMATION: Clone ID: MRT4577_1172C.1.pep
US-10-425-115-203631

Query Match 16.3%; Score 262.5; DB 17; Length 195;
Best Local Similarity 31.8%; Pred. No. 3.8e-10;
Matches 64; Conservative 46; Mismatches 78; Indels 13; Gaps 3;

QY 8 QDIQECTTOIYLYKQVQPSVAQLRSTWVDPAINLFLKMKGELETKDLQQAQNLSA 67
Db 3 QBITELKSAVRDLTKQLRPPSMQTRLLIDPAIHFEFTRLKNLVEEKKIKELQDNVAA 62
QY 68 WKFTPDSTGKIKMAKCRMLIQENELGRLSQGRIAGLEAEALQKYSEELKSSQDEL 127
Db 63 VNFPTSSKLGKMLAKCRTLQENEEIGTMASEGKIHELGMKIAVLKQNNELRNQFDVL 122
QY 128 NDFIQLDEEVEGMSSTILVLQOQLKETROQLAQYQOQOQ-SQASAPSTSRRTASEPVEQSE 187
Db 123 YKHMGGVNDVERSNELVAILQEELEAKELEVLRLKEETLSCKGCTPEAPAP 173
QY 188 ATSKDCSRLTNGPSNGSSSRQ 208
Db 174 -TSDDAG---NGQEADSDTLQ 190

RESULT 8

US-10-425-114-48881
; Sequence 48881, Application US/10425114
; Publication No. US20040034888A1

```

; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 48881
; LENGTH: 213
; TYPE: PRT
; ORGANISM: Oryza sativa nipponbare
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3474-004-D12_FLI.pap
US-10-425-114-48891

Query Match      15.9%; Score 256; DB 15; Length 213;
Best Local Similarity 31.8%; Pred. No. 1.2e-09;
Matches 55; Conservative 47; Mismatches 71; Indels 0; Gaps 0;

QY 4 ATKQDIQECTTQIQYLKQVQSPVAQLRSTWDPAINLFFLMKMGELQTKDKLEQAQN 63
DB 16 AKRQEIAELKSAVRDLTKQLRPPSMQTRLLDPAIHFEFTRLKNLVEKEKKIKELQD 75
QY 64 ELSAWKTPDSOTGKKLMAKRMILQENBELGRQLSGRIQAELAEALQKYSSEELKSS 123
DB 76 NVAAVNFTPSSKHGKMLMAKCRITQEENEEIGAMASEGKIHELGMKIAVLKTRNNELRQ 135
QY 124 QDELNDFIQIDEEVEGMSQSTILVLQQLKETQQLAQYQQQSQASAPSTR 176
DB 136 FNELYKMDGLTNDVERSNEMVAIIQDELETQDVRLKEMLAQKEATDENK 188

RESULT 9
US-10-437-963-196776
; Sequence 196776, Application US/10437963
; Publication No. US20040133343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbaruk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 196776
; LENGTH: 336
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)...(336)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_92596C.1.pap
US-10-437-963-196776

Query Match      15.9%; Score 256; DB 16; Length 336;
Best Local Similarity 31.8%; Pred. No. 2e-09;
Matches 55; Conservative 47; Mismatches 71; Indels 0; Gaps 0;

```

```

QY 4 ATKQDIQECTTQIQYLKQVQSPVAQLRSTWDPAINLFFLMKMGELQTKDKLEQAQN 63
DB 139 AKRQEIAELKSAVRDLTKQLRPPSMQTRLLDPAIHFEFTRLKNLVEKEKKIKELQD 198
QY 64 ELSAWKTPDSOTGKKLMAKRMILQENBELGRQLSGRIQAELAEALQKYSSEELKSS 123
DB 199 NVAAVNFTPSSKHGKMLMAKCRITQEENEEIGAMASEGKIHELGMKIAVLKTRNNELRQ 258
QY 124 QDELNDFIQIDEEVEGMSQSTILVLQQLKETQQLAQYQQQSQASAPSTR 176
DB 259 FNELYKMDGLTNDVERSNEMVAIIQDELETQDVRLKEMLAQKEATDENK 311

RESULT 10
US-10-310-154-593
; Sequence 593, Application US/10310154
; Publication No. US20030233670A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; APPLICANT: Chomet, Paul S.
; APPLICANT: Adams, Thomas H
; APPLICANT: Ruff, Thomas G.
; APPLICANT: Agarwal, Ameeta K.
; APPLICANT: Ahrens, Jeffrey E.
; APPLICANT: Ball, James A.
; APPLICANT: Banu, G.
; APPLICANT: Bell, Brin
; APPLICANT: Boddupalli, Raghava
; APPLICANT: Deikman, Jill
; APPLICANT: Deng, Molian
; APPLICANT: Dong, Jinzhao
; APPLICANT: Duff, Stephen M.
; APPLICANT: Galligan, Meghan M.
; APPLICANT: Hinchey, Brenda S.
; APPLICANT: Huang, Shihshieh
; APPLICANT: Johnson, G. Richard
; APPLICANT: Jung, Vincent
; APPLICANT: Kretzmer, Keith A.
; APPLICANT: Laccetti, Lucille B.
; APPLICANT: Lai, Chao-Qiang
; APPLICANT: Lee, Gary
; APPLICANT: Lin, Jie-Yi
; APPLICANT: Liu, Jingdong
; APPLICANT: Lu, Bin
; APPLICANT: Lu, Michael M.
; APPLICANT: Lund, Adrian
; APPLICANT: Madson, Linda L.
; APPLICANT: Malloy, Kathleen A.
; APPLICANT: McKiel, Christine L.
; APPLICANT: Miller, Philip W.
; APPLICANT: Padmavathi, Manchikanti
; APPLICANT: Parnell, Laurence D.
; APPLICANT: Start, William G.
; APPLICANT: Tennesen, Dan
; APPLICANT: Viciya, K.R.
; APPLICANT: Wang, Haiyuo
; APPLICANT: Xin, Zhanguo
; APPLICANT: Xu, Nanfei
; APPLICANT: Yang, Chunzhi
; APPLICANT: Zeng, Xiaoping
; APPLICANT: Zhang, Qiang
; APPLICANT: Zhao, Yajuan
; APPLICANT: Zhou, Li
; TITLE OF INVENTION: Gene Sequences and Uses Thereof in Plants
; FILE REFERENCE: 38-15(52796)B
; CURRENT APPLICATION NUMBER: US/10/310,154
; CURRENT FILING DATE: 2002-12-04
; PRIOR APPLICATION NUMBER: 60/337,358
; PRIOR FILING DATE: 2001-12-04
; NUMBER OF SEQ ID NOS: 736
; SEQ ID NO 593
; LENGTH: 352
; TYPE: PRT

```

```
; ORGANISM: Oryza sativa
US-10-310-154-593

Query Match      15.9%; Score 256; DB 14; Length 352;
Best Local Similarity 31.8%; Pred. No. 2.2e-09;
Matches 55; Conservative 47; Mismatches 71; Indels 0; Gaps 0;

QY 4 ATKGDIOECTTOIQYLKQVQPSVAQLRSTWVDPAINLFFLKMGELCTDKLEQAQN 63
DB 155 AKREQETIAELKSAVRDLKQLRPPSQWTRLLDPAIHFEFRLKULVEKEKKLEQD 214
QY 64 ELSAWKFTPDSTQTKMLMAKCRMLIQENQELGRLSQGRIAQLAEALQKYSBELKSS 123
DB 215 NVAANFTPSKKGKMLMAKCRTLQENEEIGAVASEGKIHELGMKIAVLKTRNNELRNQ 274
QY 124 QDELNDFIQLDEVEGMSQSTIQLQOLKEITQQLAQYQQOQSASAPSTR 176
DB 275 FNELYKHMGLTNDVRSNEMVAILODELTKDVELRLUKMLAQKATDENK 327

RESULT 11
US-10-425-114-55340
; Sequence 55340, Application US/10425114
; Publication No. US20040034889A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 55340
; LENGTH: 190
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: JC-GMLE01810068D05_FLI.pep
US-10-425-114-55340

Query Match      15.4%; Score 247; DB 15; Length 190;
Best Local Similarity 35.3%; Pred. No. 4.2e-09;
Matches 49; Conservative 40; Mismatches 50; Indels 0; Gaps 0;

QY 30 QLRSTWDPAINLFLKMGKLEQTKDLEQAQNELSAWKFTPDSTQTKMLMAKCRMLIQ 89
DB 8 QARLLDPVHBEFTRLNILVEEKDKVKELQNIAAVSFTPSKMGKMLMAKCRTLQE 67
QY 90 ENQELGRLSQGRIAQLAEALQKYSBELKSSQDELNDFIQLDEVEGMSQSTIQLVQ 149
DB 58 ENEEIGNQASEGKWHELGMKALQKSONSLRNQFEGKQKHMEGLTNDVRSNEMWLMQ 127
QY 150 QOLKETROQLAQYQQOQSQ 168
DB 128 EKLEEKDRIQLRKHEIQ 146

RESULT 12
US-10-739-930-9291
; Sequence 9291, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53377)B
; CURRENT APPLICATION NUMBER: US/10/739,930
; CURRENT FILING DATE: 2003-12-18
```

```
; NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO 9291
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: GLYMA-23APR03-C6342_2.p
US-10-739-930-9291

Query Match      14.8%; Score 238; DB 17; Length 299;
Best Local Similarity 36.8%; Pred. No. 3e-08;
Matches 50; Conservative 35; Mismatches 51; Indels 0; Gaps 0;

QY 4 ATKGDIOECTTOIQYLKQVQPSVAQLRSTWVDPAINLFLKMGKLEQTKDLEQAQN 63
DB 146 AKREQETIAELKSAVRDLKVLKPPSQMSRRLLDPVHBEFTRLNILVEEKDKVKELQD 205
QY 64 ELSAWKFTPDSTQTKMLMAKCRMLIQENQELGRLSQGRIAQLAEALQKYSBELKSS 123
DB 206 NIAAVSFTPSKMGKMLMAKCRTLQENEEIGNQASEGKWHELGMKALQKSONSLRSQ 265
QY 124 QDELNDFIQLDEVE 139
DB 266 FEGLOKMEGLTNDVE 281

RESULT 13
US-10-424-599-240263
; Sequence 240263, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285884
; SEQ ID NO 240263
; LENGTH: 310
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(310)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_58985C.1.pap
US-10-424-599-240263

Query Match      14.8%; Score 238; DB 15; Length 310;
Best Local Similarity 34.8%; Pred. No. 3.1e-08;
Matches 53; Conservative 42; Mismatches 56; Indels 2; Gaps 2;

QY 4 ATKGDIOECTTOIQYLK-QVQPSV-AQLRSTWVDPAINLFLKMGKLEQTKDLEQA 61
DB 146 AKREQETIAELKSAVRDLKVLKPPSQAAKEXLLLDASFHEFTRLNILVEEKDKVKEL 205
QY 62 QNELSAWKFTPDSTQTKMLMAKCRMLIQENQELGRLSQGRIAQLAEALQKYSBELK 121
DB 206 QDNIAAVSETPSKMGKMLMAKCRTLQENEEIGNQASEGKWHELGMKALQKSONSLRSQ 265
QY 122 SSOELNDFIQLDEVEGMSQSTIQLVQOLKE 154
DB 266 NQFEGLOKMEGLTNDVRSNEMWLMQEKLE 298

RESULT 14
US-10-424-599-240259
; Sequence 240259, Application US/10424599
```



```
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 240259
; LENGTH: 300
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_58981C.1.pep
US-10-424-599-240259

Query Match      14.6%; Score 234; DB 15; Length 300;
Best Local Similarity 36.0%; Pred. No. 5.5e-08;
Matches 49; Conservative 36; Mismatches 51; Indels 0; Gaps 0;

Qy      4 ATKGDIOECTQIQLKQVQPSVAQLRSTMDVPAINLFLKWKGELEOTKDKLEAOQN 63
Db      147 AKRQEIAELKSAVRDLKVLKPSMQSRLLDPAVHEEFTRLKNLVEKDKKVELQD 206

Qy      64 ELSAWKFTPDSTGKMLMAKRMILIQENQELGRLSQGRIQLEAEALQKYSSEELKSS 123
Db      207 NIAVSTFQSKVKGKMLMAKRTIQEENEEIGNQAFEGKHELMKMLALQKSONSLRSQ 266

Qy      124 QDELNDFFIIQDEEVE 139
Db      267 FEGLEKHEGLTNDVE 282

RESULT 15
US-10-425-114-42229
; Sequence 42229, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 42229
; LENGTH: 165
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700167006_FLI.pep
US-10-425-114-42229

Query Match      13.6%; Score 218.5; DB 15; Length 165;
Best Local Similarity 30.0%; Pred. No. 3e-07;
Matches 45; Conservative 41; Mismatches 63; Indels 1; Gaps 1;

Qy      44 FLAMKGELEQTKRLQAQNELSAWFTPDSTGKMLMAKRMILIQENQELGRLSQGRI 103
Db      2 FTRLKNLVEKDKKIELQDNVAANNFTPSKLGKVMKAKCTIQEENEEIGAMASEGI 61

Qy      104 AQLEAEALQKYSSEELKSSQDELNDFFIIQDEEVEGMSQSTLVLCQQLKETROQLAQY 163
Db      62 HELGMKIAVLKQNNELNQFDVLYKHMDGVNDVRSNELVAILQEELEKELEVTRLK 121
```

```
Qy      164 QQQ-SQASAPSTSRRTTASBPVEQSEATSKD 192
Db      122 EETLSQKEGTPDAPAPAPAPVEEGGAGND 151

Search completed: November 14, 2004, 12:48:45
Job time : 143 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2004, 12:26:59 ; Search time 39 Seconds
(without alignments)
542.447 Million cell updates/sec

Title: US-10-030-389-2

Perfect score: 1608

Sequence: 1 MRLATKGDIQECTIQIYL.....GSRHVQNGLDSSVNVQGSYL 319

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66319000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PGTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	357	22.2	151	2	US-08-933-750C-1
2	357	22.2	151	3	US-09-234-613-1
3	166	10.3	525	4	US-09-107-532A-5095
4	156	9.7	897	1	US-08-095-737-4
5	156	9.7	897	1	US-08-480-145-4
6	156	9.7	897	2	US-08-477-389-4
7	155	9.6	57	4	US-09-621-976-5165
8	144	9.0	461	1	US-08-186-222-2
9	144	9.0	475	4	US-09-248-796A-14568
10	144	9.0	716	3	US-09-219-983A-20
11	141	8.8	1742	4	US-09-385-962C-4
12	141	8.8	1742	4	US-09-386-959-4
13	141	8.8	3248	5	PCT-US95-16216-1
14	141	8.8	3248	5	US-09-914-259-24
15	140	8.7	1031	4	US-08-328-254-6
16	139	8.6	2482	1	US-08-328-254-6
17	139	8.6	3210	4	US-09-538-092-1154
18	137.5	8.6	703	3	US-08-910-925-4
19	137	8.5	1024	4	US-09-270-767-44973
20	136.5	8.5	717	3	US-08-910-925-1
21	134.5	8.4	1155	4	US-09-710-279-1780
22	134	8.3	645	4	US-09-919-172-41
23	131.5	8.2	896	1	US-08-095-737-2
24	131.5	8.2	896	1	US-08-480-145-2
25	131.5	8.2	896	2	US-08-477-389-2
26	131.5	8.2	896	4	US-09-538-092-1113
27	131.5	8.2	1857	4	US-09-917-254-91

Sequence 1084, Ap
Sequence 919, Appl
Sequence 3, Appli
Sequence 4, Appli
Sequence 3, Appli
Sequence 2084, A
Sequence 14401, A
Sequence 2, Appli
Sequence 2, Appli
Sequence 16333, A
Sequence 1078, Ap
Sequence 520, Appl
Sequence 17, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-08-933-750C-1
; Sequence 1, Application US/08933750C
; Patent No. 5932442
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/933,750C
FILING DATE: September 23, 1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0356 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 151 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: U937NOT01

Query Match 22.2%; Score 357; DB 3; Length 151;
Best Local Similarity 95.9%; Pred. No. 1e-21;
Matches 70; Conservative 2; Mismatches 1; Indels

149 NDLMEOQKQDKKA---VEDKKAEND--AKLKELAEQAALLESQ--KGDLLSKQADLNVLK 201

Db 202 TSLAAEQATADKADLNRC-KAEAEAEQARIREQORLAEOARQAAQAEKAEAREQAE 260
QY 167 SQASABSTRTTASEPVEQSEATSKDCSRLTNGPSNGSSSRORTSGSGFHREGNTTDDF 226
Db 261 AEAQAQASTAQSSEATSSATQSSMTSESSATQSSATERTT-----PSSSTEESTA 315
QY 227 PSSPGNGKNSNSERTGRGSGYVNOISAGYVESVDSPTGSENSLTHOSN-----DT 279
Db 316 PES--SATESSTTAPESATEESTTVPESSESTTVP--ESSSTESSTTAPPTTST 370
QY 280 DSHD-----PQEKAV-----SGKN-----RTVGRHVQNG 307
Db 371 DQSVDTGNGTSPAPTPTTPEQPKPVTAPAPSGSVNGAAVAEAYKIGTPIVWG 430
QY 308 LDSS 311
Db 431 KDPS 434

RESULT 4

US-08-095-737-4
; Sequence 4, Application US/08095737
; Patent No. 5487979
; GENERAL INFORMATION:
; APPLICANT: DiFiore, Pier P
; APPLICANT: Fazioli, Francesca
; TITLE OF INVENTION: A Substrate for the Epidermal Growth
; TITLE OF INVENTION: Factor Receptor Kinase
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: United States of America
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/095,737
; FILING DATE: 19930722
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Israel, Ned A
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH060.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 897 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-095-737-4

Query Match 9.7%; Score 156; DB 1; Length 897;
Best Local Similarity 23.7%; Pred. No. 0.00027;
Matches 79; Conservative 56; Mismatches 101; Indels 98; Gaps 16;

QY 3 LATKGODIQCETQIYLYK-QVQPSV-----AQLRSTWVDPAINFLLKMKGELEQTKDK 57
Db 356 LKEKEDTVKQRTSEVDQLQDEVOREINLQKLQAKQVQ-----ELLGELDEQKAQ 407
QY 58 LEQAQNELSAWKFTPDSTQTKKLMKACRMILQENQELGRQLSQGRITAEALALQK--- 114
Db 408 LEEQLQOE-----VRKKC-----AEEAQL-----ISSLKAEITSQESQI 440

QY 115 -KYSEELKSSQDELNDF---IIQLDEVEGMSQSTILVLOQLKETROQLAQQOQSQAS 170
Db 441 SSYEELKAREELSRLOQETAQLSESVESGKAQLEPLQOHLQESQOEISSMQRLM-- 498
QY 171 APSRTTASEPVEQSEATSKDCSRLTNGPSNGSSSRORTSGSGFHREGNTTDDFPPSP 230
Db 499 -----KDL-ETDNNQSNWSSSPQSVLVNG-----ATDYCSLS 529
QY 231 GNGKNSNSERTGRGSGYVNOISAGYVESVDSPTGSENSLTHOSN-----DT 279
Db 530 TSSETATNFNHAEGQ-----NNLES-----BPTHQESSV-RSSPEIAPSDVTDESEA 576
QY 291 VSGKGNRTVGR-----HVQNGLDSSVNVQSVL 319
Db 577 VTVAGNEKVTRFDDDKHKE--EDPFNVESSL 608

RESULT 5

US-08-480-145-4
; Sequence 4, Application US/08480145
; Patent No. 5717067
; GENERAL INFORMATION:
; APPLICANT: DiFiore, Pier P
; APPLICANT: Fazioli, Francesca
; TITLE OF INVENTION: A Substrate for the Epidermal Growth
; TITLE OF INVENTION: Factor Receptor Kinase
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: United States of America
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,145
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/095,737
; FILING DATE: 22-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Israel, Ned A
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH060.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 897 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-480-145-4

Query Match 9.7%; Score 156; DB 1; Length 897;
Best Local Similarity 23.7%; Pred. No. 0.00027;
Matches 79; Conservative 56; Mismatches 101; Indels 98; Gaps 16;

QY 3 LATKGODIQCETQIYLYK-QVQPSV-----AQLRSTWVDPAINFLLKMKGELEQTKDK 57
Db 356 LKEKEDTVKQRTSEVDQLQDEVOREINLQKLQAKQVQ-----ELLGELDEQKAQ 407
QY 58 LEQAQNELSAWKFTPDSTQTKKLMKACRMILQENQELGRQLSQGRITAEALALQK--- 114
Db 408 LEEQLQOE-----VRKKC-----AEEAQL-----ISSLKAEITSQESQI 440

QY 115 -KYSELKSSQDELNDP---IIQLDEVEGMOSTILVLOOQLKETROQLAQYQQQSOAS 170
 Db 441 SSYEBELLKAREELRLQOETAQLAESVESGKAQLEPLQCHLQESQOETISSNQMLEM-- 498
 QY 171 APSTRTTASEPVEOSEATSKDCSRLTNGPNSGSSSRQRTSGSGFHREGNTTDDPFPSP 230
 Db 499 -----KDL-ETDNNQSNWSSSPQSVLVNG-----ATDYCSLS 529
 QY 231 GNGKSSNSBERTGRGGSGYVNLASAGYVSDPTGSENLSLTHQSDNTDSSHDPPQBEKA 290
 Db 530 TSSSTANFNHAEQO-----NNLES-----EPHQESSV-RSSPEIAPSDVTDESEA 576
 QY 291 VSGKNRTVGRS-----HVQNLGSSVNVQGSVL 319
 Db 577 VTVAGNEKVTFRFDDDKHSKE--EDPFNVESSSL 608

RESULT 6
 US-08-477-389-4
 ; Sequence 4, Application US/08477389
 ; Patent No. 5872219
 ; GENERAL INFORMATION:
 ; APPLICANT: Difiore, Pier P
 ; APPLICANT: Fazioli, Francesca
 ; TITLE OF INVENTION: A Substrate for the Epidermal Growth
 ; TITLE OF INVENTION: Factor Receptor Kinase
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Knobbe, Martens, Olson & Bear
 ; STREET: 620 Newport Center Drive, Sixteenth Floor
 ; CITY: Newport Beach
 ; STATE: California
 ; COUNTRY: United States of America
 ; ZIP: 92660
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION NUMBER: US/08/477,389
 ; FILING DATE: 07-JUN-1995
 ; CLASSIFICATION: 530
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/095,737
 ; FILING DATE: 22-JUL-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Israelsen, Ned A
 ; REGISTRATION NUMBER: 29,655
 ; REFERENCE/DOCKET NUMBER: NIH060.001A
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (619) 235-8550
 ; TELEFAX: (619) 235-0176
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 897 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-477-389-4

Query Match 9.7%; Score 156; DB 2; Length 897;
 Best Local Similarity 23.7%; Pred. No. 0.00027;
 Matches 79; Conservative 56; Mismatches 101; Indels 98; Gaps 16;

QY 3 LATKGQDIQCTTQIOYLK-QVQPSV-----AQLRSTWVDPAINLFLKMKGELEQTKDK 57
 Db 356 LKEDVTQRTSEVDQLQDEQVRESINLQKQAQQQVQ-----ELLGELDEQRAQ 407
 QY 58 LEQAQNELSAWKFTPDSTQTKLMKACRMLIQENBELGRQLSGRIQAQLAEALQK--- 114
 Db 408 LERLQLE-----VRKQC-----ABEAQL-----ISSLKARITSQESQI 440

QY 115 -KYSELKSSQDELNDP---IIQLDEVEGMOSTILVLOOQLKETROQLAQYQQQSOAS 170
 Db 441 SSYEBELLKAREELRLQOETAQLAESVESGKAQLEPLQCHLQESQOETISSNQMLEM-- 498
 QY 171 APSTRTTASEPVEOSEATSKDCSRLTNGPNSGSSSRQRTSGSGFHREGNTTDDPFPSP 230
 Db 499 -----KDL-ETDNNQSNWSSSPQSVLVNG-----ATDYCSLS 529
 QY 231 GNGKSSNSBERTGRGGSGYVNLASAGYVSDPTGSENLSLTHQSDNTDSSHDPPQBEKA 290
 Db 530 TSSSTANFNHAEQO-----NNLES-----EPHQESSV-RSSPEIAPSDVTDESEA 576
 QY 291 VSGKNRTVGRS-----HVQNLGSSVNVQGSVL 319
 Db 577 VTVAGNEKVTFRFDDDKHSKE--EDPFNVESSSL 608

RESULT 7
 US-09-621-976-5165
 ; Sequence 5165, Application US/09621976
 ; Patent No. 6639063
 ; GENERAL INFORMATION:
 ; APPLICANT: Dumas Milne Edwards, J.B.
 ; APPLICANT: Jobert, S.
 ; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
 ; FILE REFERENCE: GENSET.054PR2
 ; CURRENT APPLICATION NUMBER: US/09/621,976
 ; NUMBER OF SEQ ID NOS: 19335
 ; SOFTWARE: Patent.pm
 ; SEQ ID NO 5165
 ; LENGTH: 57
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-09-621-976-5165

Query Match 9.6%; Score 155; DB 4; Length 57;
 Best Local Similarity 86.8%; Pred. No. 8e-06;
 Matches 33; Conservative 2; Mismatches 1; Indels 2; Gaps 1;

QY 26 PSVAQ--LRSTWVDPAINLFLKMKGELEQTKDKLEOA 61
 Db 20 PRLAENKLRSWVDPAINLFLKMKGELEQTKDKLEOA 57

RESULT 8
 US-08-186-222-2
 ; Sequence 2, Application US/08186222
 ; Patent No. 5559007
 ; GENERAL INFORMATION:
 ; APPLICANT: Surli, Bruno
 ; APPLICANT: Schmitz, Albert
 ; TITLE OF INVENTION: Bacterial Vectors
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: CIBA-GEIGY Corporation
 ; STREET: 7 Skyline Drive
 ; CITY: Hawthorne
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10532
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/186,222
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/672,205

FILING DATE: 19-MAR-1991
APPLICATION NUMBER: GB 9006400.7
FILING DATE: 22-MAR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Villamizar, JoAnn
REGISTRATION NUMBER: 30,598
REFERENCE/DOCKET NUMBER: 4-17994/A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914)785-7121
TELEFAX: (914)347-5769
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 461 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-186-222-2

Query Match 9.0%; Score 144; DB 1; Length 461;
Best Local Similarity 22.0%; Pred. No. 0.001;
Matches 65; Conservative 53; Mismatches 124; Indels 54; Gaps 9;

QY 2 RLATGQDIQECTTQIYKQVQPSVAQLRSTWVDPAINLFFL-----KMKGE 50
DB 88 QIATLINESIKERTKLE--AARSQVNSATNYDAVNSKSLTDVIQKVTATVSSA 145
QY 51 LEQTKDLQQAQNELSAWFTPDSTQTKY----LMAKRMILQENQELGRLSGORIAQL 106
DB 146 NKMLFQEQKEQKELK-----QKSTVKNYQFVLSQSLDSQAQELTSQAELKVATL 200
QY 107 EABLALQKYSBELKSSQDELNDFIQLDEEVEGMCSTILVLQOQLKETROQLAQQQQ 166
DB 201 -----NYQATIAQDKQAL---LDEKA-----AAKAAQBAKKAQYEAQQ 241
QY 167 SQASAPSTERTTASEPVEOSEATSKDCSLTNPGNSGSSQRTSGSGFHREGNTTDDF 226
DB 242 KEAAQQAQASTA-----ATAKAVEATSSASASSQAPQVSTST---TDNTTGNAS 288
QY 227 PSPGNGKSSNSSEERTGRGSGYVNSLSAGYSDVPTGSGNSLTHQSDNTDSS 282
DB 289 ASNSNSNSNSNSSS 344

RESULT 9
US-09-248-796A-14568
Sequence 14568, Application US/09248796A
Patent No. 6747137
GENERAL INFORMATION:
APPLICANT: Keith Weinstock et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.132
CURRENT APPLICATION NUMBER: US/09/248,796A
PRIOR FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074,725
PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/096,409
PRIOR FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 28208
SEQ ID NO 14568
LENGTH: 475
TYPE: PRT
ORGANISM: Candida albicans
US-09-248-796A-14568

Query Match 9.0%; Score 144; DB 4; Length 475;
Best Local Similarity 22.9%; Pred. No. 0.0011;
Matches 75; Conservative 53; Mismatches 132; Indels 68; Gaps 14;
QY 19 YLKQVQPSVAQLRS-----TWVDPAINLFFLQKMGELQTKDKLEQAQNELSAW 68
DB 112 YTPSSQPSGERSLSQRAPSSNDLLADPAIS-----COLSQATSDIANVNSQIKSL 164

QY 69 KETPDSTQCKKLMK---CRMILQENQELGRLSQGR-----TAQLEAEALALOKKY 116
DB 165 TSQTNLHEKTRAEKELQRLITKS--EENKQLRASYDNEVIQVEQVEANLATAKEE 223
QY 117 SEELKS-----SQDELNDFIQLDES---VEGMCSTILVLQOQLKETROQLAQQQQ--- 165
DB 224 TEALRSEASIAEAKVNSLSGELHEKQVAMESLQENSTLKEKLSLNAEIVELEKQAASK 283
QY 166 -----SQASAPSTERTTASEPVEOSEATSKDCS-RLTNPGNSGSSSQRT 210
DB 284 SQETHALNSQVAVKKSQVCVAIVKSEELKSKIAETEAASHKQLQLDLDAEERLDSNR- 342
QY 211 SGSGFHREGNTTDDFPSPGNG-----NKSNSSEERTGRGSGYVNSLSAG-YESVDS 264
DB 343 -AKELHAKSVLEQNKPTKPSAGITATIGAAAAGVAAVGGGVASVVHLSAGSTESSDS 401
QY 265 PTGSENSLTHQS--NDTSSHDPOEKA 290
DB 402 --GSRNDPVERSKEETESDNDVEETNA 427

RESULT 10
US-09-219-983A-20
Sequence 20, Application US/09219983A
Patent No. 6380159
GENERAL INFORMATION:
APPLICANT: Woliner, Mariana
APPLICANT: Lung, Oliver
APPLICANT: Tran, Khanh-Uyen
TITLE OF INVENTION: GENES FOR MALE ACCESSORY GLAND PROTEINS IN DROSOPHILA
FILE REFERENCE: 19603/1791
CURRENT APPLICATION NUMBER: US/09/219,983A
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/071,315
PRIOR FILING DATE: 1997-12-23
NUMBER OF SEQ ID NOS: 35
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 20
LENGTH: 716
TYPE: PRT
ORGANISM: Drosophila melanogaster
US-09-219-983A-20

Query Match 9.0%; Score 144; DB 3; Length 716;
Best Local Similarity 18.6%; Pred. No. 0.0019;
Matches 78; Conservative 66; Mismatches 113; Indels 162; Gaps 13;

QY 5 TKGQDIQECTTQIYKQVQVQ-PSVAQLRSTWVDPAINLFFLQKMGELQTKDKLEQAQ 63
DB 339 SKNSLSQSOSQSOBQLQDLQDLRLQLE-----QVKLEMONIRELLQKGS 385
QY 64 ELSAMKFTPDSTQTKKLMKCRMILQENQELGRLSGQRIQLEAEALQKYSLELKS 122
DB 386 ELQT-----QSDSQPRIHELYQNILQNLKEKLSYQLAKLKELED-----QKKSQAEISK 436
QY 125 SQDELNDFI----- 132
DB 437 GSNPSNLIIGLPFSEKPAFNGQPSIEPKLVQPGLPKLPFGGGLIGKPGASTGLYL 496
QY 133 -----QLDEEVEGMCSTILVLQOQLKETROQLAQQY---QQQCSQ----- 168
DB 497 SPDFNDLSYRDQFLQQLQELKHKHNLILLORRNDKKQKQNAQLLLQKQKEQAQESIN 556
QY 169 -----ASAPSTERTTASEPVE-----QSEATSKDCSLTNPGNSGSSR----- 207
DB 557 KQSSSSAGSSCTKLQDIQSTGAQSQQLQAGSTGLQTS--CGTSSASQALQRLKE 616
QY 208 -----QRTSGGFHREGNTTDDFPSSPGNGKNSNSSEERTGRGSGYVNSLS 256
DB 617 QELQRIQENDQKTSSSSSSHNSQ-----NSGSSSSSQSSQASQSAQAEAGNRNTL- 668
QY 257 AGYESVDSPTGSENSLTHQSDNTDSSHDPOEKA VSGKGNRTVSGSRHVQNGLOSSVNVQ 315

```

Db 669 -----LDQSSKTSRSP-----SRRLNRRHINRSORETHLO 703
      | : | | |
      | : | | |

```

```

RESULT 11
US-09-386-962C-4
; Sequence 4, Application US/09386962C
; Patent No. 6635473
; GENERAL INFORMATION:
; APPLICANT: FOSTER, Timothy
; TITLE OF INVENTION: POLYPEPTIDES AND POLYNUCLEOTIDES FROM COAGULASE-NEGATIVE STAPHYLOC
; FILE REFERENCE: P66335US2/BAS
; CURRENT APPLICATION NUMBER: US/09/386,962C
; CURRENT FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: 60/098,443
; PRIOR FILING DATE: 1998-08-31
; PRIOR APPLICATION NUMBER: 60/117,119
; PRIOR FILING DATE: 1999-01-25
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 1742
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-386-962C-4

```

```

Query Match      8.8%; Score 141; DB 4; Length 1742;
Best Local Similarity 24.3%; Pred. No. 0.011;
Matches 74; Conservative 48; Mismatches 124; Indels 58; Gaps 13;

Qy 41 NLFF-----LNMKGLEQTKLEQAQNELSAWKTPTDSQTGKKLMACRMILQENQE 93
Db 1 NIYPNWRYSMKRQGPINKRVDFLSNKNVAKYSIRKET-----VGTASILVGATLM 51
Qy 94 LGRQLSGRIAQLAEALALOKKYSEELKSSQDELNDFFIQLDEBVEGQSTILVLOQLK 153
Db 52 FGAADNEAKAAE---DNQLSASKEEGKGRDHNENSKLNQVDLD-NGSHS-----EK 100
Qy 154 ETRQLQLAYQQQSQASAPST---SRTTASPEVQESEATSKDCSRLTNGSPNGSSSRQRT 210
Db 101 TTNVNNA---TEVKKVEAPTSDVSKPKANEAVVTNESTK--PKTEAETVNEESIAET 154
Qy 211 SGSGFPHREGNTTDDFPSSPGNGKNSNSSEERTGRCGCGYVNLQS---AGYSRVDSPTG 267
Db 155 PKSTSTTQQ-DSTKKNPNSLKNDLNSSTTSKESKTDHEHSTKQAQMSNTKSNLNDNSPTQ 213
Qy 268 SENSLLTHQSNDTDSHSD-----PQEBKAVSGKGN--RTVGSRHVQNGL-----D 309
Db 214 SEKTSSCANNDSTNQSAPSQLDSKPSQKVTYKTFNDEPTQDVHHTTKLKTPTPSVSTD 273
Qy 310 SSVN 313
Db 274 SSVN 277

```

```

RESULT 12
US-09-386-959-4
, Sequence 4, Application US/09386959
, Patent No. 6703025
, GENERAL INFORMATION:
, APPLICANT: PATTI, Joseph M.
, APPLICANT: FOSTER, Timothy J.
, APPLICANT: HOOK, Magnus
, TITLE OF INVENTION: MULTICOMPONENT VACCINES
, FILE REFERENCE: P0633303/US/BAS
, CURRENT APPLICATION NUMBER: US/09/386,959
, CURRENT FILING DATE: 1999-08-31
, EARLIER APPLICATION NUMBER: 607098,439
, EARLIER FILING DATE: 1999-08-31
, NUMBER OF SEQ ID NOS: 65
, SOFTWARE: PatentIn Ver. 2.0
, SEQ ID NO 4
, LENGTH: 1742

```

```

; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-386-959-4

Query Match      8.8%; Score 141; DB 4; Length 1742;
Best Local Similarity 24.3%; Fred. No. 0.011;
Matches 74; Conservative 48; Mismatches 124; Indels 58; Gaps 13;

Qy 41 NLFF-----LKKMGELEQTKLEQAOQNELSAWKFTPDSTQTKLMAKCRMLIQENQE 93
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 1 NIYFNWYSMKRRQGPINKRVDFLSKNVKNYSIRKFT-----VGTASILVGATLM 51

Qy 94 LGRQLSGRTAQLEAELALQKKYSEBLKSSQDELNFIQLDDEVEGMOQTILVLOOQLK 153
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 52 FGAADNEAKAE---DNQLESASKEEQKGRDVENSKLNQVDLD-NGSHSS-----EK 100

Qy 154 ETRQQLAQYQQOQSASAPST---SRTTASEPVEQSEATSKDCSRUTNGPSNGSSSRQRT 210
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 101 TTNVNA---TEVKYEAPTISDVSPKANEAVVTHSEIK---PKTEAPTNVNEESIAET 154

Qy 211 SGSGFHREGNTTDEDDFPSSPGNKNKSNSEERTGRGGGYNNQLS---AGYESVDSPG 267
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 155 PKTSTTQQ-DSTEKNPFLKNDLINSSTTTSKESKTDHSTKQAQMSTNKNLNDTNDSPQT 213

Qy 268 SENSLTHQSNDDTSSHD-----PQEKAVSGKGN--RTVGSRHVQNGL-----D 309
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 214 SEITSSQANDSTDNQSAFSKQLDSRPSQKVYKTFNDEPTQDVEHTTTKLTPTSVSTD 273

Qy 310 SSVN 313
   |||
Db 274 SSVN 277

```

RESULT 13
 US-08-353-700-1
 ; Sequence 1, Application US/08353700
 ; Patent No. 5599919
 ; GENERAL INFORMATION:
 ; APPLICANT: YEN, TIMOTHY J.
 ; APPLICANT: RATNER, JEROME B.
 ; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
 ; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
 ; TITLE OF INVENTION: AND METHODS OF USE
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: DANN, DOREFAN, HERRELL AND SKILLMAN
 ; STREET: 1601 MARKET STREET, SUITE 720
 ; CITY: PHILADELPHIA
 ; STATE: PA
 ; COUNTRY: USA
 ; ZIP: 19103-2307
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/353,700
 ; FILING DATE: 09-DEC-1994
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: REED, JANET E.
 ; REGISTRATION NUMBER: 36,252
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (215) 563-4100
 ; TELEFAX: (215) 563-4044
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 3248 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein


```

; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HUMAN
; US-08-353-700-1

Query Match      8.8%; Score 141; DB 1; Length 3248;
Best Local Similarity 22.4%; Pred. No. 0.025;
Matches 73; Conservative 61; Mismatches 118; Indels 74; Gaps 15;

QY 12 ECTTQIQLKQVQVQSVQALRSTWVDPAINLFLKMKGELEQTQKLEQAQNELSAWKFT 71
Db 2857 KATTQI-----LEELKTKYD--NLKY-----VNQLKENERAQGM----- 2891

QY 72 PDSOTGKLMKAKRMILQENQELGRQLSQRIQAQLEAB--LALOKKYSE---ELKSSQDE 126
Db 2892 -----KLLIKSCQLBEKEILQKELSOLQAQEKQKTGVMTKVDELTTTEIKELKET 2945

QY 127 LNDFIQLDREVEGQSTILVLOOQLKETROQL-----AQYQQQSQASAPSTSRIT 178
Db 2946 LEETKKEADEYLD-KYCSLLISHEKLEKAKEMLETVAHLCSSQSKQDSRGSPLLGPVVP 3004

QY 179 ASEPVQSEATSKDCSRLTNGPSNGSSSRORTSGSGFHREGNTTDDFPSPGNGKSN 238
Db 3005 GPSPI--PSVTEK--RLSSGQNKASGKRORSSGIWENGRT-----PATPESFKSKK 3054

QY 239 SS-----BERTGRGGSGYVNLQLSAGYVSDPTGSENLSLTHOSNDTSSHDPO-- 286
Db 3055 KAVMSGIHPAEDTGTETFEPEGLPEVVKGF--ADIPTGKTSPIYLRTTMTATRTSPRLA 3112

QY 287 -EKAUS-----GKGNRTVGRHVQNG 307
Db 3113 AQKALSPSLGKXENLAESSKPTAGG 3138
```

```

RESULT 14
PCT-US95-16216-1
; Sequence 1, Application PC/TUS9516216
; GENERAL INFORMATION:
; APPLICANT: Yen, Timothy J.
; APPLICANT: Rattner, Jerome B.
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
; TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16216
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,700
; FILING DATE: 09-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
```

```

; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; PCT-US95-16216-1

Query Match      8.8%; Score 141; DB 5; Length 3248;
Best Local Similarity 22.4%; Pred. No. 0.025;
Matches 73; Conservative 61; Mismatches 118; Indels 74; Gaps 15;

QY 12 ECTTQIQLKQVQVQSVQALRSTWVDPAINLFLKMKGELEQTQKLEQAQNELSAWKFT 71
Db 2857 KATTQI-----LEELKTKYD--NLKY-----VNQLKENERAQGM----- 2891

QY 72 PDSOTGKLMKAKRMILQENQELGRQLSQRIQAQLEAB--LALOKKYSE---ELKSSQDE 126
Db 2892 -----KLLIKSCQLBEKEILQKELSOLQAQEKQKTGVMTKVDELTTTEIKELKET 2945

QY 127 LNDFIQLDREVEGQSTILVLOOQLKETROQL-----AQYQQQSQASAPSTSRIT 178
Db 2946 LEETKKEADEYLD-KYCSLLISHEKLEKAKEMLETVAHLCSSQSKQDSRGSPLLGPVVP 3004

QY 179 ASEPVQSEATSKDCSRLTNGPSNGSSSRORTSGSGFHREGNTTDDFPSPGNGKSN 238
Db 3005 GPSPI--PSVTEK--RLSSGQNKASGKRORSSGIWENGRT-----PATPESFKSKK 3054

QY 239 SS-----BERTGRGGSGYVNLQLSAGYVSDPTGSENLSLTHOSNDTSSHDPO-- 286
Db 3055 KAVMSGIHPAEDTGTETFEPEGLPEVVKGF--ADIPTGKTSPIYLRTTMTATRTSPRLA 3112

QY 287 -EKAUS-----GKGNRTVGRHVQNG 307
Db 3113 AQKALSPSLGKXENLAESSKPTAGG 3138
```

```

RESULT 15
US-09-914-259-24
; Sequence 24, Application US/09914259
; Patent No. 5495336
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/09/914,259
; CURRENT FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 1031
; TYPE: PRT
; ORGANISM: Strongylocentrotus purpuratus
; US-09-914-259-24

Query Match      8.7%; Score 140; DB 4; Length 1031;
Best Local Similarity 21.3%; Pred. No. 0.0064;
Matches 70; Conservative 50; Mismatches 125; Indels 84; Gaps 11;

QY 2 RIATKQDIOECTTQIQLKQVQSVQALRSTWVDPAINLFLKMK-----GEL 51
Db 710 QLANLRTNEKEHQWELKQVQ-----RMTLQHEKQLDYKLEKIEAEKAAKRL 763

QY 52 EOTKDKLEQAQNELSAWKFTPDSOTGKIL--MAKCRMLI-----QENQE 93
Db 764 SQQFDRREQAKQDLKLEET---VAKELQTLNLKLFVSDLNQRVKKALEGGDRDDDS 819

QY 94 LGHQLSCGRIQAQLEAELALQKXYSEELKSSQDELDNFILQDEEVEGQSTILVLOOQLK 153
Db 820 GGSQAQKQKISFLNNLEQLTKVHQLVDRNADRLCELPKRLRLRATSERVALEMSLK 879

QY 154 ETRQ----QLAQYQQQSQASAPSTSRITASEPVQSEATSKDCSRLTNGPSNGSSSROR 209
```

```

Db      ||:::      :|||:      :      :|||:      :|||:
880 ETKGAMRDRKRYQOEVDRI-----REAVROKFAKRGSSAQIA 918

QY      :      :      :      :      :      :      :
210 TSGSGFHREGNTTDDFPSSPCNGKNSSEERTGRGSGYVNLGAGYESVDSPT-GS 268

Db      :      :      :      :      :      :      :
919 KAIRAGHP-----PSPGGS-----TGIRGGYSGIRGGSPVIRPPSHGS 959

QY      :      :      :      :      :      :      :
269 ENSLTHQSDTSDSHDPQOEKAVSGKGNR 297

Db      :      :      :      :      :      :      :
960 PEPISH-NNSPEKSLNPNDNAENMEKKANK 987

```

Search completed: November 14, 2004, 12:37:14
 Job time : 41 secs